

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Holly Schnizer Examiner #: 7658 Date: _____
 Art Unit: 1653 Phone Number 30 _____ Serial Number: 09/444281
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Seq ID 35 linked to Seq ID #37 followed
 by Seq ID #35

~~Reprint~~
 Seq ID #36

BEST AVAILABLE COPY

STAFF USE ONLY

Searcher: <u>Sheppard</u>	Type of Search	Vendors and cost where applicable
Searcher Phone #: <u>308-4494</u>	NA Sequence (#) _____	STN _____
Searcher Location: _____	AA Sequence (#) _____	Dialog _____
Date Searcher Picked Up: _____	Structure (#) _____	Questel/Orbit _____
Date Completed: <u>11/17/03</u>	Bibliographic _____	Dr.Link _____
Searcher Prep & Review Time: _____	Litigation _____	Lexis/Nexis _____
Clerical Prep Time: _____	Fulltext _____	Sequence Systems _____
Online Time: _____	Patent Family _____	WWW/Internet _____
	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:01:45 : Search time 15 Seconds
(without alignments)
76,908 Million cell updates/sec

Title: US-09-444-281-36

Perfect score: 86

Sequence: 1 ILRPMWMPWRRK 12

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	81.4	144	1 JCI1222	indolicidin precursor
2	53	61.6	1173	1 VGIHHC	E2 glycoprotein pr
3	51	59.3	299	2 T12505	hypothetical prote
4	50	58.1	111	2 T29295	hypothetical prote
5	49	57.0	467	2 E89605	protein F18G5.2 [l
6	49	57.0	498	1 JT0751	ferredoxin-NADP re
7	48.5	56.4	114	2 T36208	hypothetical prote
8	48	55.8	265	2 AH0755	conserved hypochet
9	47	54.7	248	2 S23449	NADH oxidase (H202
10	47	54.7	253	2 G70715	hypothetical prote
11	47	54.7	276	2 B83161	probable short-cha
12	47	54.7	715	2 B70741	probable most prot
13	47	54.7	1411	2 T48529	hypothetical prote
14	46	53.5	728	2 T51071	related to tria pr
15	45.5	52.9	505	2 A39128	anthranilate synth
16	45	52.3	187	2 AC3353	hypothetical prote
17	45	52.3	196	2 S55483	modulator of drug
18	45	52.3	273	2 F82646	monofunctional bio
19	45	52.3	412	2 A83604	probable MFS trans
20	45	52.3	448	2 H72376	hypothetical prote
21	45	52.3	1108	2 A48508	cyclic-nucleotide
22	44	51.2	67	2 AC1954	hypothetical prote
23	44	51.2	257	2 S70177	yfire protein - Yer
24	44	51.2	353	2 AB1823	hypothetical prote
25	44	51.2	361	2 A36669	3-alpha-galactosyl
26	44	51.2	397	2 B70763	probable membrane
27	44	51.2	355	2 T38244	hypothetical prote
28	44	51.2	621	2 S37664	peplimeric polypro
29	44	51.2	630	2 S37663	peplimeric polypro

30	44	51.2	967	2 C70831	probable mmp14 pro
31	44	51.2	968	2 F70746	probable mmp12 pro
32	44	51.2	968	2 T00322	hypothetical prote
33	44	51.2	1154	1 VGIHHC	E2 glycoprotein pr
34	44	51.2	1162	1 VGIHHC	E2 glycoprotein pr
35	44	51.2	1162	2 S07421	E2 glycoprotein pr
36	44	51.2	1162	2 S14939	E2 glycoprotein pr
37	44	51.2	1162	2 S14940	E2 glycoprotein pr
38	43.5	50.6	276	2 AH0244	probable esterase
39	43.5	50.6	1529	2 A59189	ATP-binding casses
40	43	50.0	51	2 S23291	light-harvesting p
41	43	50.0	192	2 H86543	hypothetical prote
42	43	50.0	192	2 D72081	conserved hypochet
43	43	50.0	236	2 J00606	arylesterase (EC 3
44	43	50.0	250	2 A83506	probable cobalamin
45	43	50.0	278	2 T46458	hypothetical prote

ALIGNMENTS

RESULT 1
JCI1222
indolicidin precursor - bovine
N:Alternate names: antimicrobial peptide
C:Species: Bos primigenius taurus (cattle)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JCI1222; A42387; S25664
R:del Sal, G.; Storicl, P.; Schneider, C.; Romeo, D.; Zanetti, M.
Biochem. Biophys. Res. Commun. 187, 467-472, 1992
A:Title: cDNA cloning of the neutrophil bactericidal peptide indolicidin.
A:Reference number: JCI1222; MUID:92392368; PMID:1520337
A:Accession: JCI1222
A:Molecule type: mRNA
A:Residues: 1-144 <SAL>
A:Cross-references: EMBL:X67340; NID:9462; PTDN:CA47755.1; PID:9463
A:Experimental source: EMBL: bone marrow
R:Seisted, M.E.; Novotny, M.J.; Morris, W.L.; Tang, Y.Q.; Smith, W.; Cullor, J.S.
J. Biol. Chem. 267, 4292-4295, 1992
A:Title: indolicidin, a novel bactericidal tridecapeptide amide from neutrophils.
A:Reference number: A42387; MUID:92165771; PMID:1537821
A:Accession: A42387
A:Molecule type: protein
A:Residues: 131-143 <SEL>
A:Experimental source: neutrophils
A:Note: sequence extracted from NCBI backbone (NCBIP:83840)
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-143/Product: indolicidin #status experimental <MAT>
F:143/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 81.4% Score 70; DB 1; Length 144;
Best Local Similarity 88.9% Pred. No. 0.014; 0; Indels 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 0;

Oy 3 RWPMPWRR 11
Db 135 KWPMPWRR 143

RESULT 2
E2 glycoprotein precursor - human coronavirus (strain 229E)
N:Alternate names: peplomer glycoprotein; S glycoprotein; spike glycoprotein
C:Species: human coronavirus
A:Note: host Homo sapiens (man)
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C:Accession: A34766; S05460
R:Raabe, T.; Schelle-Prinz, B.; Siddell, S.G.
J. Gen. Virol. 71, 1065-1073, 1990

```

A:Cross-references: EMBL:U41557; PIDN:AAA83303.1; CESP:C50F7.8
C:Genetics:
A:Gene: CESP:C50F7.8

Query Match          58.1%; Score 50; DB 2; Length 111;
Best Local Similarity 54.5%; Pred. No. 4.2;
Matches      6; Conservative      2; Mismatches      3; Indels      0; Gaps      0;

QY      1 ILRMPMPMRR 11
          :: ||||| |
Db      12 VMMPMPMPCGR 22

RESULT 5
E89605
Protein F18G5.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C:Accession: E89605
R:Anonymous; The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: E89605
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-467 <STO>
A:Cross-references: ST:chr_X; PIDN:AAA81082.1; PID:g1055093; GSPDB:GN00028; CESP:F18G
C:Genetics:

```

```

Query Match          57.0%; Score 49; DB 2; Length 467;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 WPWPWW 9
      ||||| 1
Db      201 WPWPWW 206

RESULT 6
JT0751
ferredoxin-NADP reductase (EC 1.18.1.2), long form precursor - bovine
N:Alternate names: adrenodoxin reductase
C:Species: Bos primigenius taurus (cattle)
C:Date: 14-Jul-1994 #sequence_revision 18-Oct-1996 #text_change 03-Jun-2002
C:Accession: J070751, J00079; J0390; S03558; P50003; A29604; S52100
R:Rakata, Y.; Sagarra, Y.; Kono, A.; Sekimizu, K.; Horiiuchi, T.
Biol. Pharm. Bull. 16, 1200-1206, 1993
A:Title: Gene structure of bovine adrenodoxin reductase.
A:Reference number: JT0751; M01D:94I77140; PMID:8130767
A:Accession: J070751
A:Molecule type: DNA
A:Residues: 1-498 <TRAK>
A:Cross-references: GB:D83475; NTD:g1199916; P1DN:BAAl1921.1; PID:g4521308
A:Experimental source: adrenal cortex
A:Note: the authors translated the codon GTC for residue 205 as Gly
R:Sagarra, Y.; Takata, Y.; Miyata, T.; Hara, T.; Horiiuchi, T.
J. Biochem. 102, 1333-1336, 1987
A:Title: Cloning and sequence analysis of adrenodoxin reductase cDNA from bovine adrenal
A:Reference number: JT0079; M01D:86198050; PMID:3448086
A:Accession: JT0079
A:Molecule type: mRNA
A:Residues: 1-204, 211-498 <SAG>
A:Cross-references: GB:D00211; NTD:g217433; P1DN:BA00150.1; PID:g217434
A:Note: the deduced sequence is partially confirmed by amino acid sequencing of 15
R:Sagarra, Y.
submitted to DBJ, September 1989
A:Reference number: J0390
A:Contents: revision, insertion of residues 205-210

```

A:Accession: J50390
 A:Molecule type: mRNA
 A:Residues: 56-498 <SA2>
 R:Hanukoglu, I.; Gutfinger, T.
 Eur. J. Biochem. 180, 479-484, 1989
 A:Title: cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites in
 A:Reference number: S03558; MUID:91170752; PMID:2924777
 A:Accession: S03558
 A:Molecule type: mRNA
 A:Residues: 155-204, 211-498 <HAN>
 A:Cross-references: EMBL:X13736; MID:965; PIDN:CAA32002.1; PID:9833776
 A:Note: 40S-Ser was also found
 R:Hamamoto, I.; Kurokouchi, K.; Tanaka, S.; Ichikawa, Y.
 Biochim. Biophys. Acta 953, 207-213, 1988
 A:Title: Adrenodoxin-binding peptide of NADPH-adrenodoxin reductase.
 A:Reference number: P50003; MUID:88184054; PMID:3355838
 A:Accession: P50003
 A:Molecule type: Protein
 A:Residues: 33-41, 'S', 43-62, 260-283, 'TW', 496-498 <HAM>
 A:Note: a cyanogen bromide peptide binds to adrenodoxin
 R:Nonaka, Y.; Murakami, H.; Yabusaki, Y.; Kuramitsu, S.; Kagamiyama, H.; Yamano, T.; Oka
 Biochem. Biophys. Res. Commun. 145, 1239-1247, 1987
 A:Title: Molecular cloning and sequence analysis of full-length cDNA for mRNA of adrenod
 A:Reference number: A29604; MUID:87270695; PMID:3038094
 A:Accession: A29604
 A:Molecule type: mRNA
 A:Residues: 1-76, 'R', 78-80, 'VWIALTPRSRL', 95-123, 'RVYRLF', 129-204, 211-273, 'R', 275-322,
 A:Cross-references: GB:M17029; MID:9162628; PIDN:AAA30362.1; PID:9162629
 A:Experimental source: adrenal cortex
 R:Warburton, R.J.; Seybert, D.W.
 Biochim. Biophys. Acta 1246, 39-46, 1995
 A:Title: Structural and functional characterization of bovine adrenodoxin reductase by 1
 A:Reference number: S52100; MUID:95110846; PMID:7811729
 A:Accession: S52100
 A:Status: Preliminary
 A:Molecule type: Protein
 A:Residues: 'X', 34-41, 'X', 43-48, 'X', 50-51, 304-306, 'X', 308-309, 'X', 311-326 <WAR>
 C:Comment: Ferredoxin-NADP+ reductase is localized in the matrix of adrenal cortex mito
 eridoxin-NADP+ reductase, adrenodoxin and two forms of cytochrome P-450.
 C:Genetics:
 A:Introns: 27/1; 59/3; 91/3; 132/3; 170/3; 204/3; 246/3; 275/1; 341/3; 399/1; 456/1
 C:Function:
 A:Description: catalyzes the reversible reduction of NADP+ by reduced ferredoxin or red
 C:Superfamily: human ferredoxin-NADP+ reductase
 C:Keywords: alternative splicing; flavoprotein; mitochondrion; monomer; NADP; oxidoreduc
 F:1-32/Domain: transit peptide (mitochondrion) #status predicted <SIG>
 F:33-498/Product: ferredoxin-NADP+ reductase, long form #status predicted <MAT>
 F:33-204, 211-498/Product: ferredoxin-NADP+ reductase, short form #status experimental <
 F:40-70/Region: beta-alpha-beta FAD nucleotide-binding fold
 F:180-190/Region: NADP binding #status predicted
 F:281/Binding site: substrate (lys) #status experimental

Query Match 57.0%; Score 49; DB 1; Length 498;
 Best Local Similarity 83.3%; Pred. No. 23;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WPMWPM 9
 Db 6 WPMWPM 11

RESULT 7
 T36208
 hypothetical protein SCE36.09 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36208
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL data library, May 1999
 A:Reference number: 221601
 A:Accession: T36208
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-114 <OLI>
 A:Cross-references: EMBL:AL049763; PIDN:CAB42078.1; GSPDB:GN00070; SCOEDB:SCE36.09
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCE36.09

Query Match 56.4%; Score 48.5; DB 2; Length 114;
 Best Local Similarity 80.0%; Pred. No. 6.7;
 Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 3 RW-PWMPWR 11
 Db 103 RWPRWPMWR 112

RESULT 8
 AH0755
 conserved_hypothetical protein STY2208 [imported] - Salmonella enterica subsp. enteri
 C:Species: Salmonella enterica subsp. enterica serovar Typh
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AH0755
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 , S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AH0755
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-265 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD05747.1; PID:916503239; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY2208

Query Match 55.8%; Score 48; DB 2; Length 265;
 Best Local Similarity 31.6%; Pred. No. 17;
 Matches 6; Conservative 3; Mismatches 0; Indels 10; Gaps 1;

Qy 1 ILRMPW-----WPM 9
 Db 1 MIKMPWKAQETQNEDEWPM 19

RESULT 9
 S23449
 NADH oxidase (H2O2-forming) (EC 1.6.-.-) - Thermus aquaticus
 C:Species: Thermus aquaticus
 C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 23-Mar-1993
 C:Accession: S23449; S24556
 R:Park, H.U.; Kreutzger, R.; Reiser, C.O.A.; Sprinzl, M.
 Eur. J. Biochem. 205, 875-879, 1992
 A:Title: Molecular cloning and nucleotide sequence of the gene encoding a H(2)O(2)-fo
 A:Reference number: S23449; MUID:92249331; PMID:1577004
 A:Accession: S23449
 A:Molecule type: DNA
 A:Residues: 1-248 <PAR>
 A:Cross-references: EMBL:X60110
 A:Accession: S24556
 A:Molecule type: protein
 A:Residues: 1-32 <PARL>
 C:Genetics:
 A:Gene: nox
 C:Keywords: NAD; oxidoreductase
 F:1-248/Product: NADH oxidase (H2O2-forming) #status experimental <MAT>

Query Match 54.7%; Score 47; DB 2; Length 248;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PMWPM 9

Db 179 PWMPW 183

RESULT 10

hypothetical protein Rv0945 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: G70715

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70715

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-253 <COL>

A:Cross-references: GB:279700; GB:AL123456; NID:93261628; PIDN:CA802005.1; PID:91524217

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv0945

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F:8-190/Domain: short-chain alcohol dehydrogenase homology <SABH>

Query Match 54.7%; Score 47; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PWMPW 9
| | | | |
Db 230 PWMPW 234

RESULT 11

probable short-chain dehydrogenase PA3883 [Imported] - Pseudomonas aeruginosa (strain PA

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: B83161

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Micoquchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

Lox, S.; Olson, M.V.

Nature 406, 959-964, 2000

A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: B83161

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <STO>

A:Cross-references: GB:AE004805; GB:AE004091; NID:9950055; PIDN:AA607270.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3883

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 54.7%; Score 47; DB 2; Length 276;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 RWPMPWRRK 12
| | | | |
Db 197 RSPWMPWRQ 206

RESULT 12

B70741

probable moe protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70741

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70741

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-715 <COL>

A:Cross-references: GB:275555; GB:AL123456; NID:93261608; PIDN:CAA99988.1; PID:e25035

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: moe

Query Match 54.7%; Score 47; DB 2; Length 715;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 RWPMPWRRK 11
| | | | |
Db 65 RWAYYPWRR 73

RESULT 13

hypothetical protein T22P22.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48529

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Banco

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24490

A:Accession: T48529

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1411 <BEV>

A:Cross-references: EMBL:AL163814

A:Experimental source: cultivar Columbia; BAC clone T22P22

C:Genetics:

A:Map position: 5

A:Insertions: 281/2; 320/1; 389/3; 429/3; 473/3; 515/3; 534/2; 567/3; 602/1; 669/1; 776/

A>Note: T22P22.90

Query Match 54.7%; Score 47; DB 2; Length 1411;
Best Local Similarity 63.6%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 LRMPMPWRRK 12
| | | | |
Db 1013 LAWSQWRRK 1023

RESULT 14

related to trfA protein [Imported] - Neurospora crassa

N:Alternate names: protein B2A19.50

C:Species: Neurospora crassa

C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: T51071

R:Schulte, U.; Align, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, July 2000

A:Reference number: Z25286

A:Accession: T51071

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-728 <SCH>

A:Cross-references: EMBL:AL390092; GSPDB:GN00116; NCSP:B2A19.50

A:Experimental source: BAC clone B2A19; strain OR74A

C:Genetics:

A:Gene: NCSP:B2A19.50

A;Map position: 6
A;Introns: 26/1; 119/2

Query Match	53.58;	Score 46;	DB 2;	Length 728;
Best Local Similarity	58.38;	Pred. No. 81;		
Matches	7;	Conservative	1;	Mismatches 4;
			Indels	0;
			Gaps	0;

QY	1	ILRWPPWPRRK	12
		:	
Db	11	ILGWPLWWSRR	22

RESULT 15

A39128

anthranilate synthase (EC 4.1.3.27) component I [validated] - *Pseudomonas syringae* pv. *5*
N:Alternate names: anthranilate synthase alpha chain
S:Species: *Pseudomonas syringae* pv. *savastanoi*
C:Classification: *Pseudomonas syringae* pv. *savastanoi*

C: Species: *Pseudomonas syringae* pv. *savastanoi*
N/Alterna: *Alterna* *synthase* *alpha* *chain*

C;species: Pseudomonas syringae pv. savastanoi

C;	Date:	27-Nov-1991	#sequence_revision	27-Nov-1991	#text_change	17-Mar-2000
----	-------	-------------	--------------------	-------------	--------------	-------------

C;Accession: A39128

R;da Costa, E.; Silva, O.; Kosuge, T.

J. Bacteriol. 173, 463-471, 1991

A;Title: Molecular characterization and expression analysis of the anthranilate synthase

A;Reference number: A39128; MUID:91100331; PMID:1987141

A;Accession: A39128

A;Status: preliminary

A: Molecule type: DNA

```

//: molecule type. DNA
A:Residues: 1-505 <DACC>

```

```

m/residues: 1 000 <DACC>
A:Cross-references: GB:M55911

```

C:Genetics:

C;GeneLics:

A;Gene: LIPE
C:Comb:lev: heterotetramor: two common types associated with

C; complex: hetero

c;Function: <ANT>

A;Description: EC 4.1.3.27 [valid]

A; pathway: tryptop

A;Note: 11rst step

```
C;Function: <COM1>
```

A;Description: EC 4.1.3.27 [validated, MUID:91100331]

A: Note: expression of type seems to be independent

C:Superfamily: anthranilate synthase component I

Query Match	52.98;	Score 45.5;	DB 2;	Length 505;
-------------	--------	-------------	-------	-------------

QY 1 ILRW-----PWWPWRK 12

Search completed: January 15, 2003, 18:04:27
Job time : 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 17:48:25 : Search time 11 Seconds

(without alignments)
45.247 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRMPMPMPRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	81.4	144	1 INDC_BOVIN	P33046 bos taurus
2	53	61.6	1173	1 VGL2_CVB22	P15423 human coron
3	49	57.0	492	1 ADRO_BOVIN	P08165 bos taurus
4	47	54.7	253	1 Y945_MYCTU	P71564 mycobacteri
5	47	54.7	715	1 YD55_MYCTU	Q11025 mycobacteri
6	45.5	52.9	505	1 TRPE_PSESS	P21889 pseudomonas
7	45	52.3	196	1 YAO5_SCHPO	Q09677 schizosacch
8	45	52.3	1108	1 CN3B_RAT	Q63085 rattus norv
9	44	51.2	361	1 FUT3_HUMAN	P21217 homo sapien
10	44	51.2	372	1 FUT3_HUMAN	Q19058 pan troglod
11	44	51.2	397	1 MML6_MYCTU	Q10773 mycobacteri
12	44	51.2	535	1 YDW6_SCHPO	Q13312 schizosacch
13	44	51.2	967	1 MML4_MYCTU	Q53735 mycobacteri
14	44	51.2	968	1 MML2_MYCTU	Q11171 mycobacteri
15	44	51.2	984	1 SX13_MOUSE	Q04891 mus musculu
16	44	51.2	1154	1 VGL2_IBVD2	P12722 avian infec
17	44	51.2	1162	1 VGL2_IBV	P11223 avian infec
18	44	51.2	1162	1 VGL2_IBV	P12650 avian infec
19	44	51.2	1162	1 VGL2_IBV	P12651 avian infec
20	44	51.2	1163	1 VGL2_IBV	P05135 avian infec
21	43.5	50.6	276	1 RCCL_RHOP	Q83005 rhodospheud
22	43.5	50.6	236	1 ABC2_HUMAN	Q98327 homo sapien
23	43	50.0	51	1 LHR2_RCTHA	R80105 ectothiorho
24	43	50.0	711	1 MMLA_STRCO	Q5302 streptomyc
25	43	50.0	958	1 MML1_MYCTU	P95211 mycobacteri
26	43	50.0	1112	1 CN3B_HUMAN	Q13370 homo sapien
27	43	50.0	1225	1 VGL2_CVPR8	P27655 porcine res
28	43	50.0	1225	1 VGL2_CVPR8	P24413 porcine res
29	43	50.0	1235	1 VGL2_CVPR8	P11225 murine coro
30	43	50.0	1324	1 VGL2_CVMA5	P11224 murine coro
31	43	50.0	1353	1 VGL2_CVHOC	P36334 human coron
32	43	50.0	1363	1 VGL2_CVBF	P25190 bovine coro
33	43	50.0	1363	1 VGL2_CVBL9	P25191 bovine coro

ALIGNMENTS

RESULT 1	ID	INDC_BOVIN	STANDARD:	PRT:	144 AA.
AC	P33046;	INDC_BOVIN			
DT	01-OCT-1993 (Rel. 27, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Indolicidin precursor.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone marrow;				
RA	MEDLINE=92392368; PubMed=1520337;				
RA	del Sal G., Storici P., Schneider C., Romeo D., Zanetti M.;				
RL	"CDNA cloning of the neutrophil bactericidal peptide indolicidin.";				
RT	Biochem. Biophys. Res. Commun. 187:467-472(1992).				
RN	[2]				
RP	SEQUENCE OF 131-143.				
RC	TISSUE=Neutrophils;				
RX	MEDLINE=92165771; PubMed=1537821;				
RA	Selsted M.E., Novotny M.J., Morris W.L., Tang Y.-Q., Smith W.;				
RA	Cullor J.S.;				
RT	"Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils.";				
RL	J. Biol. Chem. 267:4292-4295(1992).				
CC	-1- FUNCTION: POTENT MICROBICIDAL ACTIVITY, ACTIVE AGAINST STAPHYLOCOCCUS AUREUS AND ESCHERICHIA COLI.				
CC	-1- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.				
CC	-1- PTM: ELASTASE MIGHT BE RESPONSIBLE FOR ITS MATURATION.				
CC	-1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: X67340; CAA47755.1; -				
DR	PIR: JCI222; JCI222.				
DR	PIR: A42387; A42387.				
DR	InterPro: IPR001894; Cathelicidin.				
DR	Pfam: PF00666; Cathelicidins; 1.				
DR	ProDom: PD001838; Cathelicidins; 1.				
DR	PROSITE: PS00946; CATHELICIDINS_1; 1.				
DR	PROSITE: PS00947; CATHELICIDINS_2; 1.				
KW	Antibiotic; Amidation; Signal.				
FT	SIGNAL	1	29		POTENTIAL.
FT	PROPEP	30	130		
FT	PEPTIDE	131	143		INDOLICIDIN.
FT	MOD_RES	30	30		PYRROLIDONE CARBOXYLIC ACID (BY

Query Match	Similarity	Score	DB	Length
Best local	62.5%	Pred. No. 5.6		
Matches	5	Conservative	2	Mismatches
			1	Indels
			0	Gaps
Oy	2 LRMPWPM 9			
Db	1112 IKMPWVM 1119			
<p>RESULT 3</p> <p>ADRO_BOVIN STANDARD: PRT: 492 AA.</p> <p>AC P08165: G95KNB; (Rel. 08, Created)</p> <p>AD 01-AUG-1988 (Rel. 36, Last sequence update)</p> <p>DT 15-JUL-1998 (Rel. 41, Last annotation update)</p> <p>DT 15-JUN-2002 (Rel. 41, Last annotation update)</p> <p>DE NADPH:adrenodoxin oxidoreductase, mitochondrial precursor (EC 1.18.1.2) (Adrenodoxin reductase) (AR) (Ferredoxin-NADP(+) reductase).</p> <p>DE FDXR OR ADXR.</p> <p>GN Bos taurus (Bovine).</p> <p>OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bovinae; Bos.</p> <p>OC NCBI_TaxID=9913;</p> <p>OX [1]</p> <p>RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.</p> <p>RA MEDLINE=94177140; PubMed=8130767;</p> <p>RA Takata Y., Sagara Y., Kono A., Sekimizu K., Horiuchi T.;</p> <p>RT "Gene structure of bovine adrenodoxin reductase.";</p> <p>RL Biol. Pharm. Bull. 16:1200-1206(1993).</p> <p>RN [2]</p> <p>RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.</p> <p>RA MEDLINE=8198050; PubMed=3448086;</p> <p>RA Sagera Y., Takata Y., Miyata T., Hara T., Horiuchi T.;</p> <p>RT "Cloning and sequence analysis of adrenodoxin reductase cDNA from bovine adrenal cortex.";</p> <p>RL J. Biochem. 102:1333-1336(1987).</p> <p>RN [3]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RA MEDLINE=87270696; PubMed=3038094;</p> <p>RA Nonaka Y., Murakami H., Yabusaki Y., Kuramitsu S., Kagamiyama H.,</p> <p>RA Yamano T., Okamoto M.;</p> <p>RT "Molecular cloning and sequence analysis of full-length cDNA for mRNA of adrenodoxin oxidoreductase from bovine adrenal cortex.";</p> <p>RL Biochem. Biophys. Res. Commun. 145:1239-1247(1987).</p> <p>RN [4]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC TISSUE=Adrenal cortex;</p> <p>RX MEDLINE=8910752; PubMed=2924777;</p> <p>RA Hanuoglu I., Gultinger T.;</p> <p>RT "cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites in oxidoreductases.";</p>				

```

RN RL Eur. J. Biochem. 180:479-484(1989).
RN [5]
RN SEQUENCE OF N-TERMINUS, AND PARTIAL SEQUENCE.
RP RC TISSUE-Adrenal cortex;
RC MEDLINE=88082777; PubMed=3691502;
RA Hanukoglu I., Gutfinger T., Hanlu M., Shively J.E.;
RT "Isolation of a cDNA for adrenodoxin reductase (ferrodoxin-NADP+
RT reductase). Implications for mitochondrial cytochrome P-450 systems.";
RL Eur. J. Biochem. 165:449-455(1987).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 33-49Z.
RP RC TISSUE-Adrenal gland;
RC MEDLINE=99299392; PubMed=10369776;
RA Ziegler G.A., Vornheim C., Hanukoglu I., Schulz G.E.;
RT "The structure of adrenodoxin reductase of mitochondrial P450 systems:
RT electron transfer for steroid biosynthesis.";
RN RL J. Mol. Biol. 289:981-990(1999).
RN [7]
RN X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RP RC MEDLINE=20455764; PubMed=10996235;
RA Ziegler G.A., Schulz G.E.;
RT "Crystal structures of adrenodoxin reductase in complex with NADP+ and
RT NADPH suggesting a mechanism for the electron transfer of an enzyme
RT family.";
RN RL Biochemistry 39:10986-10995(2000).
RN [8]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF COMPLEX WITH ADRENODOXIN.
RP RC MEDLINE=21264735; PubMed=11053423;
RA Mueller J.J., Lapko A., Bourenkov G., Ruckpaul K., Heinemann U.;
RT "Adrenodoxin reductase-adrenodoxin complex structure suggests electron
RT transfer path in steroid biosynthesis.";
RN RL J. Biol. Chem. 276:2786-2789(2001).
RN [9]
RN FUNCTION: SERVES AS THE FIRST ELECTRON TRANSFER PROTEIN IN ALL THE
RN MITOCHONDRIAL P450 SYSTEMS. INCLUDING: CHOLESTEROL SIDE CHAIN
RN CLEAVAGE IN ALL STEROIDGENIC TISSUES, STEROID 11-BETA
RN HYDROXYLATION IN THE ADRENAL CORTEX, 25-OH-VITAMIN D3-24
RN HYDROXYLATION IN THE KIDNEY, AND STEROL C-27 HYDROXYLATION IN THE
RN LIVER.
RN [10]
RN CATALYTIC ACTIVITY: Reduced adrenodoxin + NADP(+) = oxidized
RN adrenodoxin + NADPH.
RN [11]
RN COFACTOR: FAD.
RN [12]
RN PATHWAY: CHOLESTEROL SIDE-CHAIN-CLEAVAGE SYSTEM.
RN [13]
RN SUBUNIT: MONOMER.
RN [14]
RN SUBCELLULAR LOCATION: Mitochondrial matrix.
RN [15]
RN ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM (SHOWN HERE) AND A
RN LONG FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE LONG FORM
RN REPRESENTS 10-20% OF ALL ADRENODOXIN REDUCTASE MRNA. AND SEEMS TO
RN BE INACTIVE.
RN [16]
RN THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on ways
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement. (see http://www.isb-sib.ch/announcement/).
RN or send an email to license@isb-sib.ch).
RN [17]
RN EMBL: D83475; BAAL1921.1; -.
RN EMBL: D83472; BAAL1921.1; JOINED.
RN EMBL: D83473; BAAL1921.1; JOINED.
RN EMBL: D83474; BAAL1921.1; JOINED.
RN EMBL: M17029; AAA03052.1; -.
RN EMBL: D00211; AAA00150.1; -.
RN EMBL: X13736; CAA32002.1; -.
RN PIR: A29604; A29604.
RN PIR: JS0390; JS0390.
RN PIR: S03558; S03558.
RN PIR: J70751; J70751.
RN PDB: 1CJC; 12-APR-99.
RN PDB: 1E1L; 02-JUN-00.
RN PDB: 1E1K; 24-SEP-00.
RN PDB: 1E1M; 24-SEP-00.
RN PDB: 1E1N; 24-SEP-00.

```

```

DR   PDB: 1E6E; 09-AUG-01.
DR   InterPro: IPR000759; Adrndx_reductase.
DR   PRINTS: PR00419; ADXRDPASE.
KW   Electron transport; Oxidoreductase; Flavoprotein; NADP; FAD;
KW   Mitochondrion; Transil peptide; Alternative splicing; 3D-structure.
FT   TRANSIT      1      32      MITOCHONDRION.
FT   CHAIN         33      492      NADPH:ADRENODOXIN OXIDOREDUCTASE.
FT   VARSPIC       204      204      E -> EVILLO (IN LONG ISOFORM).
FT   CONFLICT      77      77      G -> R (IN REF. 3).
FT   CONFLICT      81      94      FGVAHPDEKAVNVI -> VWLALTPPSRMLL (IN REF. 3).
FT   CONFLICT      124     128      ODAWH -> RYRLTL (IN REF. 3).
FT   CONFLICT      268     268      K -> R (IN REF. 3).
FT   CONFLICT      317     318      PS -> RL (IN REF. 3).
FT   CONFLICT      323     333      RAAGIRLAVTR -> ARRSAMQSP (IN REF. 3).
FT   CONFLICT      341     352      TRAVPTGPDVDEL -> HPGSAMWCGGP (IN REF. 3).
SO   SEQUENCE      492 AA; 54338 MW; E68F6F5D18F53131 CRC64;

Query Match      57.0%; Score 49; DB 1; Length 492;
Best Local Similarity 83.3%; Pred. NO. 8.3;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 WPMWPM 9
        11111
Db       6 WRWMPM 11

RESULT 4
Y945_MYCTU      STANDARD;      PRT;      253 AA.
ID   Y945_MYCTU
AC   P71564
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative oxiforeductase RV0945 (EC 1.-.-.-).
GN   RV0945 OR MFO971 OR MTCY10D7.29C.
OS   Mycobacterium tuberculosis
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteriidae;
OC   Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=968295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RT   Nature 393:537-544(1998).
   [2]
RN   SEQUENCE FROM N.A.
RP   STRAIN=CDC 1551 / Oshkosh;
RC   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolony J.F., Nelson W.C., Umayam L.A., Ermolova M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
   (SQR) FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z79700; CAB02005.1;
DR EMBL; AE006982; AAK45219.1;
DR TIGR; MT0971;
DR TubercList; RV0945;
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short.1.
DR PROSITE; PS00061; ADH_SHORT.1.
DR Hypothetical protein; Oxidoreductase; Complete proteome.
FW ACT SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 253 AA; 27138 MW; BAD93720842DA12 CRC64;

Query Match 54.7%; Score 47; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 PWM 9
Db 230 PWM 234

RESULT 5
YD55 MYCTU STANDARD; PRT; 715 AA.
AC 011025;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV1355C.
GN RV1355C OR MT1398 OR MTCY02B10.19C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwin M.L., Hart D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z75555; CAA99988.1;

```

```

DR EMBL; AE007012; AAK45661.1; ALT_INIT.
DR TIGR; MT1398;
DR TubercList; RV1355C;
DR InterPro; IPR000594; Thif_domain.
DR Pfam; PF00899; Thif.1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 715 AA; 78181 MW; 455495248A56041C CRC64;

Query Match 54.7%; Score 47; DB 1; Length 715;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 RMPWPMRR 11
Db 65 RMPWPMRR 73

RESULT 6
TRPE_PSESS STANDARD; PRT; 505 AA.
ID TRPE_PSESS
AC P21689;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Anthranilate synthase component I (EC 4.1.3.27).
GN TRPE.
OS Pseudomonas syringae (pv. savastanoi).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=29438;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9110031; PubMed=1987141;
RA da Costa e Silva O., Kosuge T.;
RT "Molecular characterization and expression analysis of the
RT anthranilate synthase gene of Pseudomonas syringae subsp.
RT savastanoi."
RL J. Bacteriol. 173:463-471(1991).
CC -1- CATALYTIC ACTIVITY: Chorismate + L-glutamine = anthranilate +
CC pyruvate + L-glutamate.
CC -1- PATHWAY: Tryptophan biosynthesis; first step.
CC -1- SUBUNIT: Tetramer of two components I and two components II (by
CC SIMILARITY).
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC GLUTAMINE AMINOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55911; AAA26016.1;
CC PIR; A39128; A39128.
DR HSSP; Q06128; IQDL.
DR InterPro; IPR005256; Anth_synthI.
DR InterPro; IPR000350; Chorismate_bind.
DR Pfam; PF00425; chorismate_bind.1.
DR PRINTS; PR00095; ANTSNTHASEI.
DR PRODOM; PD000779; Chorismate_bind.1.
DR TIGRfams; TIGR00564; trpe_most.1.
DR Tryptophan biosynthesis; Lyase.
SQ SEQUENCE 505 AA; 56084 MW; A38E8193131f6BB CRC64;

Query Match 52.9%; Score 45.5; DB 1; Length 505;
Best Local Similarity 28.0%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 3; Indels 13; Gaps 1;

```

OY 1 ILRW-----PMWPMRRK 12
Db 467 VLEMEETLNKRRAMVGSAMWPMRR 491

RESULT 7
ID YA05_SCHPO STANDARD; PRT; 196 AA.
AC 009677;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C5H10.05C in chromosome 1.
GN SPAC5H10.05C.
OS Schizosaccharomyces pombe (fission yeast)
OC Eukaryota, Fungi, Ascomycota, Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodard J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mosl D., Hubert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Calbert F., Ares S.U., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forstburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussey D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: STRONG, TO BACTERIAL MODULATOR OF DRUG ACTIVITY B
(MDAB).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49811; CA89955.1.
DR InterPro: IPR003680; NADHdh_2.
DR Pfam: PF02525; NADHdh_2; 1.
KW Hypothetical protein.
SO SEQUENCE 196 AA; 22104 MW; 436764DA9E26074C CRC64;

Query Match 52.3%; Score 45; DB 1; Length 196;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 2; Indels 4; Gaps 2;
OY 1 ILRWPMW---PMWRRK 12
Db 63 IYQWPGWMMGTPWKLK 78

RESULT 8
ID CN3B_RAT STANDARD; PRT; 1108 AA.
AC 063085;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GMP-inhibited 3',5'-cyclic phosphodiesterase B (EC 3.1.4.17) (Cyclic
DE GMP-inhibited phosphodiesterase B) (CGI-PDE B) (CGIPDE1).
GN PDE3B.
OS Rattus norvegicus (Rat).
OC Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi;
OC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Adipose tissue;
RX MEDLINE=9336761; PubMed=8395509;
RA Taira M., Hockman S.C., Calvo J.C., Taira M., Belfrage P.,
RA Manganiello V.C.;
RT "Molecular cloning of the rat adipocyte hormone-sensitive cyclic GMP-
RT inhibited cyclic nucleotide phosphodiesterase."
RL J. Biol. Chem. 268:18573-18579(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN FAT METABOLISM.
CC -1- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC nucleoside 5'-phosphate.
CC -1- ENZYME REGULATION: INHIBITED BY CGMP.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- TISSUE SPECIFICITY: ABUNDANT IN ADIPOSE TISSUES.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z22867; CA80489.1.
DR InterPro: IPR003607; ME_Plipase_HDC.
DR InterPro: IPR002073; PDEase.
DR Pfam: PF00223; PDEase; 1.
DR SMART: SM00471; Hdc; 1.
DR PROSITE: PS00126; PDEASE_1; 1.
KW Hydrolyase; GMP; Membrane.
FT DOMAIN 16 22 POLY-PRO.
FT DOMAIN 99 102 POLY-ALA.
FT DOMAIN 175 179 POLY-ALA.
FT DOMAIN 1007 1021 POLY-ASP.
FT DOMAIN 1068 1071 POLY-GLU.
FT DOMAIN 1101 1104 POLY-GLU.
SO SEQUENCE 1108 AA; 123105 MW; C9B5078C7D3ADD6D CRC64;

Query Match 52.3%; Score 45; DB 1; Length 1108;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 4 WPMWPMRR 11
Db 164 WQWMSWLR 171

RESULT 9
ID FUT3_HUMAN STANDARD; PRT; 361 AA.
AC P21217; Q99448; Q99449;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Galactoside 3(4)-L-fucosyltransferase (EC 2.4.1.65) (Blood group Lewis
DE ii).
GN FUT3 OR LE OR FT3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE FROM N.A.
RX MEDLINE=91032981; PubMed=1977660;
RA Kokowska-Iatallio J.F., Larsen R.D., Nair R.P., Lowe J.B.;
RT "A cloned human cDNA determines expression of a mouse stage-specific
RT embryonic antigen and the Lewis blood group
RT alpha(1,3/1,4)fucosyltransferase.";
RL Genes Dev. 4:1268-1303(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95378269; PubMed=7650030;
RA Cameron H.S., Szczepaniak D., Weston W.;
RT "Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
RT genes in normal tissues. Alternative splicing, polyadenylation, and
RT isoforms.";
RL J. Biol. Chem. 270:20112-20122(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Squamous cell carcinoma;
RA Rahim I., Scmidt L.R., Wahl D., Drayson E., Maslanik W.,
RA Stranahan P.L., Pettijohn D.E.;
RT "Isolation and expression of human alpha (1,3/1,4)
RT fucosyltransferase.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANT LE(-) MET-105.
RX MEDLINE=94059067; PubMed=8240322;
RA Elngren A., Rydberg L., Larsson G.;
RT "Genotypic heterogeneity among Lewis negative individuals.";
RL Biochem. Biophys. Res. Commun. 196:515-520(1993).
RN [5]
RP MEDLINE=94059082; PubMed=8240337;
RA Nishihara S., Yazawa S., Iwasaki H., Nakazato M., Kudo T., Ando T.,
RA Nishimatsu H.;
RT "Alpha (1,3/1,4)fucosyltransferase (FUCT-TII) gene is inactivated by
RT a single amino acid substitution in Lewis histo-blood type negative
RT individuals.";
RL Biochem. Biophys. Res. Commun. 196:624-631(1993).
RN [6]
RP MEDLINE=94033579; PubMed=8219240;
RX Koda Y., Kimura H., Mekada E.;
RT "Analysis of Lewis fucosyltransferase genes from the human gastric
RT mucosa of Lewis-positive and -negative individuals.";
RL Blood 82:2915-2919(1993).
RN [7]
RP MEDLINE=95050753; PubMed=7961897;
RA Nishihara S., Nishimatsu H., Iwasaki H., Yazawa S., Akamatsu S.,
RA Ando T., Sano T., Nishimatsu I.;
RT "Molecular genetic analysis of the human Lewis histo-blood group
RT system.";
RL J. Biol. Chem. 269:29271-29278(1994).
RN [9]
RP VARIANTS LE(-) ARG-20; ARG-68; MET-105 AND LYS-356.
RN

FX	MEDLINE=96243526; PubMed=8801770; Rydberg L., Larson G.?
RA	Elmgren A., Boerjeson C., Svensson L.,
RT	"DNA sequencing and screening for point mutations in the human Lewis
RT	RT 'FUT3' gene enables molecular genotyping of the human Lewis blood
RT	group system.";
RL	Vox Sang. 70:97-103(1996).
RN	[10]
RP	VARIANTS LE(-) ARG-68 AND MET-105.
RX	MEDLINE=97413801; PubMed=9268337.
RA	Elmgren A., Mollicone R., Costache M., Boerjeson C., Oriol R.,
RA	Harrington J., Larson G.?
RT	"Significance of individual point mutations, T202C and C314T, in the
RT	human Lewis 'FUT3' gene for expression of Lewis antigens by the human
RT	alpha1,3/1,4-fucosyltransferase, Fuc-TII.";
RL	J Biol. Chem. 272:21994-21998(1997).
RN	[11]
RP	VARIANTS LE(+) K-102 AND A-124, AND VARIANTS LE(-) N-162; R-223 AND
RP	M-270.
RX	MEDLINE=98366989; PubMed=9703429;
RA	Pang H., Liu Y., Koda Y., Soejima M., Jia J., Schlaphoff T.,
RA	du Toit E.D., Kimura H.;
RT	"Five novel missense mutations of the Lewis gene 'FUT3' in African
RT	'Xhosa' and Caucasian populations in South Africa."
RL	Hum. Genet. 102:675-680(1998).
CC	-I- FUNCTION: MAY CATALYZE ALPHA-1,3 AND ALPHA-1,4 GLYCOSIDIC LINKAGES
CC	INVOLVED IN THE EXPRESSION OF YLM-2, LEWIS A, LEWIS B, STAYL
CC	LEWIS X AND LEWIS X/SSBA-1 ANTIGENS. MAY BE INVOLVED IN BLOOD
CC	GROUP LEWIS DETERMINATION: LEWIS-POSITIVE (LE(+)) INDIVIDUALS
CC	HAVE AN ACTIVE ENZYME WHILE LEWIS-NEGATIVE (LE(-)) INDIVIDUALS
CC	HAVE AN INACTIVE ENZYME.
CC	-I- CATALYTIC ACTIVITY: GDP-L-fucose + 1,3-beta-D-galactosyl-N-
CC	acetyl-D-glucosaminyl-R = GDP + 1,3-beta-D-galactosyl-(alpha-
CC	1,4-L-fucosyl)-N-acetyl-D-glucosaminyl-R.
CC	-I- PATHWAY: Glycosylation.
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC	FORM IN TRANS CISTERNAE OF GOLGI.
CC	-I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN STOMACH, COLON, SMALL
CC	INTESTINE, LONG AND KIDNEY AND TO A LESSER EXTENT IN SALIVARY
CC	GLAND, BLADDER, UTERUS AND LIVER.
CC	-I- MISCELLANEOUS: ALSO ACTS ON THE CORRESPONDING 1,4-GALACTOSYL
CC	DERIVATIVE, FORMING 1,3-L-FUCOSYL LINKS.
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL distribution
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: X53578; CAA37641.1; -
DR	EMBL: U27328; AAC50187.1; -
DR	EMBL: U27326; AAC50185.1; -
DR	EMBL: U27327; AAC50186.1; -
DR	EMBL: D69324; BAA13941.1; -
DR	EMBL: D69325; BAA13942.1; -
DR	EMBL: AF131913; AAD33514.1; -
DR	PIR: A36669; A36669.
DR	Genew: HGNC:4014; FUT3.
DR	MIM: 111100; -
DR	InterPro: IPR001503; GT_10.
DR	Pfam: PF008652; Glyco.transf.10; 1.
KW	Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW	Signal-anchor; Golgi stack; Polymorphism; Blood group antigen.
FT	DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 35 361 LIGNAL, CATALYTIC (POTENTIAL).
FT	CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (PROBABLE).
FT	CARBOHYD 185 185 L-LINKED (GLCNAC. . .) (PROBABLE).
FT	VARIANT 20 20 L -> R (IN LE(-)).
FT	VARIANT 20 20 F/TIG-VAR_003426.
FT	W -> R (IN LE(-)).
FT	VARIANT 68 68

```

FT FT VARIANT 102 102 /FTId=VAR_007959.
FT FT VARIANT 105 105 O -> K (IN LE(+)).
FT FT VARIANT 105 105 /FTId=VAR_007960.
FT FT VARIANT 124 124 T -> M (IN LE(-)).
FT FT VARIANT 124 124 /FTId=VAR_003427.
FT FT VARIANT 162 162 S -> A (IN LE(+)).
FT FT VARIANT 162 162 /FTId=VAR_007961.
FT FT VARIANT 170 170 D -> N (IN LE(-)).
FT FT VARIANT 170 170 /FTId=VAR_007962.
FT FT VARIANT 223 223 G -> S (IN LE(-)).
FT FT VARIANT 223 223 /FTId=VAR_003428.
FT FT VARIANT 270 270 G -> R (IN LE(-)).
FT FT VARIANT 270 270 /FTId=VAR_007963.
FT FT VARIANT 336 336 V -> M (IN LE(-)).
FT FT VARIANT 336 336 /FTId=VAR_007964.
FT FT VARIANT 356 356 D -> A (IN LE(-)).
FT FT VARIANT 356 356 /FTId=VAR_003429.
FT FT VARIANT 356 356 I -> K (IN LE(-)).
FT FT SEQUENCE 361 AA; 42117 MW; BF4398044F19C84 CRC64;
SO SEQUENCE 361 AA; 42117 MW; BF4398044F19C84 CRC64;

Query Match 51.2%; Score 44; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 PWMPPRR 11
Db 9 PWMPPRR 15

RESULT 10
FUT3_PANTR
ID FUT3_PANTR STANDARD; PRT; 372 AA.
AC 019058;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactoside 3(4)-L-fucosyltransferase (EC 2.4.1.65) (Blood group Lewis
DE alpha-4-fucosyltransferase) (Lewis FT) (Fucosyltransferase 3) (FUCT-
DE IIT) (Alpha-3/4-fucosyltransferase).
GN FUT3.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98037800; PubMed=9368041;
RA Costache M., Apoll P.-A., Calileau A., Elmgren A., Larson G.,
RA Henry S., Blancher A., Iordachescu D., Ortol R., Mollicone R.,
RT "Evolution of fucosyltransferase genes in vertebrates.";
RL J. Biol. Chem. 272:29721-29728(1997).
CC -1- FUNCTION: MAY CATALYZE ALPHA-1,3 AND ALPHA-1,4 GLYCOSIDIC LINKAGES
CC INVOLVED IN THE EXPRESSION OF SIALYL LEWIS X AND LEWIS X/SSA-1
CC ANTIGENS. IT MAY BE INVOLVED IN BLOOD GROUP LEWIS DETERMINATION
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: GDP-L-fucose + 1,3-beta-D-galactosyl-N-
CC acetyl-D-glucosaminyl-R = GDP + 1,3-beta-D-galactosyl-(alpha-
CC 1,4-L-fucosyl)-N-acetyl-D-glucosaminyl-R.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI (BY SIMILARITY).
CC -1- POLYMORPHISM: THERE ARE TWO ALLELES (A AND B). ALLELE A HAS ARG-
CC 162 AND VAL-304. ALLELE B HAS GLY-162 AND MET-304.
CC -1- MISCELLANEOUS: ALSO ACTS ON THE CORRESPONDING 1,4-GALACTOSYL
CC DERIVATIVE, FORMING 1,3-L-FUCOSYL LINKS.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y14033; CAA74360.1; -
CC InterPro: IPR001503; GT_10.
CC Pfam: PF00852; Glyco_transf.10; 1.
CC Transferrase; Glycosyltransferase; glycoprotein; Transmembrane;
CC Signal anchor; Golgi stack; Polymorphism.
CC DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 1 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT DOMAIN 15 34 (POTENTIAL).
CC FT DOMAIN 35 372 LUMENAL, CATALYTIC (POTENTIAL).
CC FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT VARIANT 162 162 R -> G (IN ALLELE B).
CC FT VARIANT 304 304 V -> M (IN ALLELE B).
CC FT SEQUENCE 372 AA; 43233 MW; 649C8F8BCA7BD74C CRC64;

Query Match 51.2%; Score 44; DB 1; Length 372;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 PWMPPRR 11
Db 9 PWMPPRR 15

RESULT 11
MML6_MYCTU
ID MML6_MYCTU STANDARD; PRT; 397 AA.
AC 010773;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein mmlp6.
GN MML6 OR RV1557 OR MT1608 OR MTCY48.08C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295967; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlind T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MMLP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: Z74020; CA98334.1; -
 DR EMBL: AE007027; AAK4875.1; -
 DR TIGR: MT1608; -
 DR TubercuList: Rv1557; -
 DR InterPro: IPR004869; MMP_L.
 DR Pfam: PF01176; MMP_L.
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 190 210 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 SQ SEQUENCE 397 AA; 42421 MW; 678DC86E24472BF4 CRC64;
 Query Match 51.2%; Score 44; DB 1; Length 397;
 Best Local Similarity 75.0%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 RWPMPMR 10
 DB 351 RWPMPMR 358
 RESULT 12
 YDM6_SCHPO STANDARD; PRT; 535 AA.
 ID YDM6_SCHPO
 AC 013912;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C23C11.06c in chromosome I.
 GN SPAC3C11.06c.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Schizosaccharomycetes; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; Pubmed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnietz B.,
 RA Weljens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borgez P., Langer I., Beck A., Lehmach H., Reinhardt R., Pohl T.M.,
 RA Borzym K., Zimmermann W., Weiler H., Wambutt R., Purnelle B.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore C., Hunt S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Kravetia J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: Z98559; CAB1159.1; -
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 55 75 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 201 221 POTENTIAL.
 FT TRANSMEM 346 366 POTENTIAL.
 SQ SEQUENCE 535 AA; 60124 MW; A6AE149AA2929E2 CRC64;
 Query Match 51.2%; Score 44; DB 1; Length 535;
 Best Local Similarity 50.0%; Pred. No. 40;
 Matches 6; Conservative 1; Mismatches 1; Indels 4; Gaps 1;
 QY 4 WFW---WPMR 11
 DB 183 WMSWSPSTWPMR 194
 RESULT 13
 MML4_MYCTU STANDARD; PRT; 967 AA.
 ID MML4_MYCTU
 AC 053735;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative membrane protein mmpL4.
 GN MML4 OR RV0450C OR MT0466 OR MTY037.14C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; Pubmed=96342230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Yernool M.D., Salzberg S.L.,
 RA Bishai W.;
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE MMP_L FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AL021932; CAI17407.1; -
DR EMBL: AE006949; AAK44689.1; -
DR TIGR: MT0466; -
DR TubercuList: RV0450c; -
DR InterPro: IPR004707; ActII.
DR InterPro: IPR004869; Mmpl.
DR Pfam: PF03176; Mmpl; 2.
DR TIGRfams: TIGR00833; actII; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
FT TRANSMEM 769 789 POTENTIAL.
FT TRANSMEM 793 813 POTENTIAL.
FT TRANSMEM 821 841 POTENTIAL.
FT TRANSMEM 875 895 POTENTIAL.
FT TRANSMEM 896 916 POTENTIAL.
SQ SEQUENCE 967 AA; 105234 MW; 6301014031480484 CRC64;

Query Match 51.28; Score 44; DB 1; Length 967;
Best Local Similarity 75.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 RMPWMPWR 10
|| ||| |
Db 930 RMPWMPWR 937

RESULT 14
MML2_MYCTU STANDARD; PRT; 968 AA.
AC 011171;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein mmpL2.
GN MmpL2 OR RV0507 OR MT0528 OR MTCY20G9.34.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigleier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Kell A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Omayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE Mmpl FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z77162; CAB00933.1; -
DR EMBL: AE006953; AAK44751.1; -
DR TIGR: MT0528; -
DR TubercuList: RV0507; -
DR InterPro: IPR004707; ActII.
DR InterPro: IPR004869; Mmpl.
DR Pfam: PF03176; Mmpl; 2.
DR TIGRfams: TIGR00833; actII; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 378 398 POTENTIAL.
FT TRANSMEM 763 783 POTENTIAL.
FT TRANSMEM 787 807 POTENTIAL.
FT TRANSMEM 815 835 POTENTIAL.
FT TRANSMEM 866 886 POTENTIAL.
FT TRANSMEM 891 911 POTENTIAL.
FT TRANSMEM 926 946 POTENTIAL.
FT CONFLICT 426 426 R -> H (IN REF. 2).
FT CONFLICT 656 656 E -> A (IN REF. 2).
SQ SEQUENCE 968 AA; 106201 MW; B68AE9B78164EDC0 CRC64;

Query Match 51.28; Score 44; DB 1; Length 968;
Best Local Similarity 75.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 RMPWMPWR 10
|| ||| |
Db 924 RMPWMPWR 931

RESULT 15
SX13_MOUSE STANDARD; PRT; 984 AA.
AC 004891;
DT 01-JUN-1994 (Rel. 29, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SOX-13 protein.
GN SOX13 OR SOX-13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION (ISOFORM 1).
RC TISSUE-Embryo;
RX MEDLINE=98083175; PubMed=9421502;
RA Rose J., Korver W., Oving E., Wilson A., Wagenaar G., Markman M.,
RA Lamers W., Clevers H.;
RT "High expression of the HMG box factor sox-13 in arterial walls during
RT embryonic development.";
RL Nucleic Acids Res. 26:469-476(1998).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION (ISOFORM 2).
RC TISSUE-Embryo;
RX MEDLINE=98201614; PubMed=9524265;

```

RA Kido S., Hiraoka Y., Ogawa M., Sakai Y., Yoshimura Y., Aiso S.;
RT "Cloning and characterization of mouse mSox13 cDNA.";
RL Gene 208:201-206(1998).
[3]
RN SEQUENCE OF 405-460 FROM N.A.
RP MEDLINE-93181275; PubMed-8441686;
RX Wright E.M., Snopek B., Koopman P.;
RT "Seven new members of the Sox gene family expressed during mouse
development."; Res. 21:744-744(1993).
RL Nucleic Acids Res. 21:744-744(1993).
CC -1- FUNCTION: BINDS TO THE SEQUENCE 5'-AACAAAT-3'.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC -1- PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, HIGH LEVELS OF EXPRESSION ARE
CC FOUND IN THE ARTERIAL WALLS AT 13.5 DAYS POST COITUM (DPC). LOW
CC LEVELS ARE FOUND IN THE INNER EAR AT 13.5 DPC AND IN SOME CELLS IN
CC THE THYMUS AT 16.5 DPC. EXPRESSED IN THE TRACHEAL EPITHELIUM BELOW
CC THE VOCAL CORD AND IN THE HAIR FOLLICLES AT 18 DPC.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AJ000740; CAA04278.1; -
DR EMBL; AB006329; BAA25786.1; -
DR EMBL; Z18962; CAA79487.1; -
DR PIR; S30241; S30241.
DR HSSP; Q05066; 1HRX.
DR MGD; MGT:98361; Sox13.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR KW DNA-binding; Nuclear protein; Alternative splicing.
FT DOMAIN 159 195
FT DNA BIND 397 465
FT VARSPPLIC 495 519
FT VARSPPLIC 603 609
FT VARSPPLIC 610 984
FT CONFLICT 35 35
FT CONFLICT 41 42
FT CONFLICT 195 195
SQ SEQUENCE 984 AA; 108897 MW; 7F5506EDAEB98C5 CRC64;
Query Match 51.28; Score 44; DB 1; Length 984;
Best Local Similarity 42.98; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
OY 4 WPMW-----PW 9
DB 686 WPMWTKLAEGFSPW 699

```

Search completed: January 15, 2003, 18:03:28
 Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:00:30 : Search time 29 seconds
(without alignments)
85.261 Million cell updates/sec

Title: US-09-444-281-36
Perfect score: 86
Sequence: 1 ILRWPMPWRRK 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	66.3	723	12	Q9DUC4
2	53	61.6	746	12	Q9JH31
3	53	61.6	1173	12	Q990M4
4	53	61.6	1173	12	Q990M4
5	53	61.6	1173	12	Q990M2
6	53	61.6	1173	12	Q990M1
7	53	61.6	1383	12	Q84712
8	53	61.6	1383	12	Q91AV1
9	53	61.6	1386	12	Q8Q098
10	52	60.5	1245	3	Q9Y7V5
11	51	59.3	298	17	Q82U59
12	51	59.3	299	4	Q9Y4N1
13	51	59.3	504	2	P96143
14	50	58.1	141	11	Q9CZAI
15	50	58.1	327	10	Q9AUN3
16	50	58.1	735	12	Q9DUC9

17	49	57.0	49	12	Q9DTR0	Q9dtr0 tt virus. o
18	49	57.0	226	4	Q961L9	Q961l9 homo sapien
19	49	57.0	226	4	Q9BSG8	Q9bsg8 homo sapien
20	49	57.0	467	5	Q19573	Q19573 caenorhabdi
21	49	57.0	606	16	Q988W4	Q988w4 rhizobium l
22	49	57.0	748	12	Q9DTR1	Q9dtr1 tt virus. o
23	49	57.0	750	12	Q91D04	Q91d04 tt virus. o
24	48.5	56.4	114	16	Q9X8C2	Q9x8c2 streptomyce
25	48	55.8	265	16	Q82NS5	Q82ns5 salmonella
26	48	55.8	265	16	Q825Q0	Q825q0 salmonella
27	48	55.8	540	2	Q07504	Q07504 bacillus me
28	47	54.7	92	12	Q8V7E2	Q8v7e2 tt virus. o
29	47	54.7	165	10	Q9SNM3	Q9snm3 oryza sativ
30	47	54.7	276	16	Q9HXC9	Q9hxc9 pseudomonas
31	47	54.7	734	12	Q8V711	Q8v711 tt virus. o
32	47	54.7	1411	10	Q9LYG0	Q9lyg0 arabidopsis
33	46	53.5	154	2	Q9R6J3	Q9r6j3 agrobacteri
34	46	53.5	257	17	Q8TWM9	Q8twm9 methanopyru
35	46	53.5	521	10	Q94EF3	Q94ef3 oryza sativ
36	45.5	52.9	175	12	Q91RD8	Q91rd8 tt virus. o
37	45	52.3	134	3	Q96UD2	Q96ud2 neurospora
38	45	52.3	150	10	Q8S697	Q8s697 oryza sativ
39	45	52.3	159	16	Q9KZT3	Q9kzt3 streptomyce
40	45	52.3	187	16	Q8YH19	Q8yh19 bruceella me
41	45	52.3	273	16	Q9PCR3	Q9pcr3 xyella fas
42	45	52.3	342	4	Q96BE4	Q96be4 homo sapien
43	45	52.3	412	16	Q916F7	Q916f7 pseudomonas
44	45	52.3	423	2	Q24742	Q24742 bacteroides
45	45	52.3	443	10	Q9S751	Q9s751 oryza sativ

ALIGNMENTS

RESULT 1
ID Q9DUC4 PRELIMINARY: PRT: 723 AA.
AC Q9DUC4: 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBL_TaxID=68867;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RA Okamoto H.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RX MEDLINE=20534983; PubMed=11080484;
RA Okamoto H.; Nishizawa T.; Tawara A.; Peng Y.; Takahashi M.;
RA Kishimoto J.; Tanaka T.; Miyakawa Y.; Mayumi M.;
RT "Species-specific TT viruses in humans and nonhuman primates and their
phylogenetic relatedness.";
RL Virology 277:368-378(2000).
DR EMBL: AB041959; BAB19313.1;
DR InterPro: IPR001563; Serine_carpept.
DR InterPro: IPR004219; TTVirus_unk.
DR Pfam: PF02956; TT_ORF1; 1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
SQ SEQUENCE 723 AA; 85393 MW; 232D003098766344 CRC64;

Query Match 66.3%; Score 57; DB 12; Length 723;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 PWMRR 11
Db 2 PWMRR 8

```

RESULT 2
09JH31          PRELIMINARY;      PRT;      746 AA.
ID 09JH31
AC 09JH31
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses: ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
GN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUN02;
RA Okamoto H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JUN02;
RX MEDLINE=20456801; PubMed=11003468;
RA Urita M., Okamoto H., Nishizawa T., Tawara A., Takahashi M.,
RA Iizuka H., Miyakawa Y., Mayumi M.;
RT "The entire nucleotide sequences of two distinct TT virus (TTV)
RT isolates (JUN01 and JUN02) remotely related to the original TTV
RT isolates."
RL Arch. Virol. 145:1543-1559(2000).
DR EMBL: AB028669; BAA94878.1; -.
DR InterPro: IPR004219; TTVirus_Unk.
DR Pfam: PF02956; TT_ORF1.1.
SQ SEQUENCE 746 AA; 88561 MW; E0B22953AE764E3E CRC64;

Query Match
Best Local Similarity 61.6%; Score 53; DB 12; Length 746;
Pred. No. 28;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 LRPMWMPWRK 12
DB 1 MAMGWMWRRR 11

RESULT 3
0990M4          PRELIMINARY;      PRT;      1173 AA.
ID 0990M4
AC 0990M4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Spike glycoprotein.
GN S.
OS Human coronavirus (strain 229E).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11137;
GN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=229E;
RA Bonavia A., Holmes K.V.;
RT "Viral and cellular changes in a human cell line persistently infected
RT with human coronavirus HCoV-229E."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF344186; AAK32188.1; -.
DR InterPro: IPR002551; Corona_S1.
DR InterPro: IPR002552; Corona_S2.
DR Pfam: PF01600; Corona_S1.1.
DR Pfam: PF01601; Corona_S2.1.
SQ SEQUENCE 1173 AA; 128669 MW; ABC6E0A75EBBD8A4 CRC64;

Query Match
Best Local Similarity 61.6%; Score 53; DB 12; Length 1173;
Pred. No. 42;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 2 LRPMWMPW 9
DB 1112 IKPMWVW 1119

RESULT 4
0990M3          PRELIMINARY;      PRT;      1173 AA.
ID 0990M3
AC 0990M3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Spike glycoprotein.
GN S.
OS Human coronavirus (strain 229E).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11137;
GN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=229E;
RA Bonavia A., Holmes K.V.;
RT "Viral and cellular changes in a human cell line persistently infected
RT with human coronavirus HCoV-229E."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF344187; AAK32189.1; -.
DR InterPro: IPR002551; Corona_S1.
DR InterPro: IPR002552; Corona_S2.
DR Pfam: PF01600; Corona_S1.1.
DR Pfam: PF01601; Corona_S2.1.
SQ SEQUENCE 1173 AA; 128683 MW; 9E2368160082A81A CRC64;

Query Match
Best Local Similarity 61.6%; Score 53; DB 12; Length 1173;
Pred. No. 42;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LRPMWMPW 9
DB 1112 IKPMWVW 1119

RESULT 5
0990M2          PRELIMINARY;      PRT;      1173 AA.
ID 0990M2
AC 0990M2;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Spike glycoprotein.
GN S.
OS Human coronavirus (strain 229E).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11137;
GN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=229E;
RA Bonavia A., Holmes K.V.;
RT "Viral and cellular changes in a human cell line persistently infected
RT with human coronavirus HCoV-229E."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF344188; AAK32190.1; -.
DR InterPro: IPR002551; Corona_S1.
DR InterPro: IPR002552; Corona_S2.
DR Pfam: PF01600; Corona_S1.1.
DR Pfam: PF01601; Corona_S2.1.
SQ SEQUENCE 1173 AA; 128653 MW; 8B658FCBBD1842DA CRC64;

Query Match
Best Local Similarity 61.6%; Score 53; DB 12; Length 1173;
Pred. No. 42;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

Db 1112 IKMPMWVW 1119

RESULT 6

0990M1 PRELIMINARY; PRT; 1173 AA.

AC 0990M1; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE Spike glycoprotein.
 GN 5.
 OS Human coronavirus (strain 229E).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=11137;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=229E;
 RA Bonavia A., Holmes K.V.;
 RT "Viral and cellular changes in a human cell line persistently infected with human coronavirus HCoV-229E."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF344189; AAK32191.1; -
 DR InterPro: IPR002551; Corona_S1.
 DR InterPro: IPR002552; Corona_S2.
 DR Pfam: PF01600; Corona_S1.1.
 DR Pfam: PF01601; Corona_S2.1.
 SQ SEQUENCE 1173 AA; 128760 MW; B73A165A6270152A CRC64;

Query Match 61.6%; Score 53; DB 12; Length 1173;
 Best Local Similarity 62.5%; Pred. No. 42;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LRMPMPW 9

Db 1112 IKMPMWVW 1119

RESULT 7

084712 PRELIMINARY; PRT; 1383 AA.

AC 084712; 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Spike protein.
 OS Porcine epidemic diarrhea virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=28295;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRI/87;
 RX MEDLINE=94231173; PubMed=8176382;
 RA Duarte M., Laude H.;
 RT "Sequence of the spike protein of the porcine epidemic diarrhoea virus."
 RL J. Gen. Virol. 75:1195-1200(1994).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRI/87;
 RX MEDLINE=93389433; PubMed=8397280;
 RA Bridgen A., Duarte M., Tobler K., Laude H., Ackermann M.;
 RT "Sequence determination of the nucleocapsid protein gene of the porcine epidemic diarrhoea virus confirms that this virus is a transmissible gastroenteritis virus."
 RL J. Gen. Virol. 74:1795-1804(1993).
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRI/87;

RX MEDLINE=94120721; PubMed=8291230;
 RA Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M.,
 RT "Sequence analysis of the porcine epidemic diarrhoea virus genome between the nucleocapsid and spike protein genes reveals a polymorphism."
 RL Virology 198:466-476(1994).
 DR EMBL: Z25483; CAA80971.1; -
 DR InterPro: IPR002551; Corona_S1.
 DR InterPro: IPR002552; Corona_S2.
 DR Pfam: PF01600; Corona_S1.1.
 DR Pfam: PF01601; Corona_S2.1.
 FT CONFLICT 422 422 Y -> N (IN REF. 1).
 SQ SEQUENCE 1383 AA; 151405 MW; 741C84DD3BDC4D CRC64;

Query Match 61.6%; Score 53; DB 12; Length 1383;
 Best Local Similarity 62.5%; Pred. No. 48;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LRMPMPW 9

Db 1321 IKMPMWVW 1328

RESULT 8

091AV1 PRELIMINARY; PRT; 1383 AA.

AC 091AV1; 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Spike protein.
 OS Porcine epidemic diarrhea virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=28295;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV777;
 RX MEDLINE=93389433; PubMed=8397280;
 RA Bridgen A., Duarte M., Tobler K., Laude H., Ackermann M.;
 RT "Sequence determination of the nucleocapsid protein gene of the porcine epidemic diarrhoea virus confirms that this virus is a transmissible gastroenteritis virus."
 RL J. Gen. Virol. 74:1795-1804(1993).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV777;
 RX MEDLINE=94120721; PubMed=8291230;
 RA Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M.,
 RT "Sequence analysis of the porcine epidemic diarrhoea virus genome between the nucleocapsid and spike protein genes reveals a polymorphic ORF."
 RL Virology 198:466-476(1994).
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV777;
 RX MEDLINE=96112302; PubMed=8830538;
 RA Tobler K., Ackermann M.;
 RT "PBDV leader sequence and junction sites."
 RL Adv. Exp. Med. Biol. 380:541-542(1995).
 RN 14
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV777;
 RX MEDLINE=98455678; PubMed=9782358;
 RA Bridgen A., Kocherhans R., Tobler K., Carvajal A., Ackermann M.;
 RT "Further analysis of the genome of porcine epidemic diarrhoea virus."
 RL Adv. Exp. Med. Biol. 440:781-786(1998).
 RN 15
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV777;
 RX MEDLINE=98455678; PubMed=9782358;
 RA Kocherhans R., Bridgen A., Ackermann M., Tobler K.;

RT "The complete genome sequence of porcine epidemic diarrhea
 RT coronavirus."
 RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF353511; AK38656.1; -
 DR InterPro: IPR002551; Corona_SI.
 DR InterPro: IPR002552; Corona_S2.
 DR Pfam: PF01600; Corona_SI; 1.
 DR Pfam: PF01601; Corona_S2; 1.
 SQ SEQUENCE 1383 AA; 151352 MW; 02E5E5E5435876D CRC64;

Query Match 61.6%; Score 53; DB 12; Length 1383;
 Best Local Similarity 62.5%; Pred. No. 48;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LRPMWMPW 9
 Db 1321 IKPMWVW 1328

RESULT 9

ID 080098 PRELIMINARY; PRT; 1386 AA.
 AC 080098;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Spike protein.
 GN SPK1.
 OS porcine epidemic diarrhea virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=28295;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kang T.-J., Lam Y.-Y., Jang Y.-S., Kwon T.-H., Kim D.-H., Yang M.-S.;
 RT "Spike Protein gene of Korea Porcine Epidemic Diarrhea Virus."
 RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF500215; AAM19716.1; -
 SQ SEQUENCE 1386 AA; 151853 MW; 11F98BC2AA0526B CRC64;

Query Match 61.6%; Score 53; DB 12; Length 1386;
 Best Local Similarity 62.5%; Pred. No. 48;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 LRPMWMPW 9
 Db 1324 IKPMWVW 1331

RESULT 10

ID 09Y7V5 PRELIMINARY; PRT; 1245 AA.
 AC 09Y7V5;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Candidospor surface protein.
 GN CMPL.
 OS Trichoderma harzianum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Trichoderma.
 OX NCBI_TaxID=5544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 32173;
 RX MEDLINE=99343881; PubMed=10413618;
 RA Puyesky M., Benhamou N., Ponce Noyola P., Bauw G., Ziv T.,
 RA Van Montagu M., Herrera Estrella A., Horvitz B.A.;
 RT "Developmental regulation of cmpl, a gene encoding a multidomain
 RT Candidospor surface protein of Trichoderma."
 RL Fungal Genet. Biol. 27:88-99(1999).
 DR EMBL: AJ133651; CAB40845.1; -
 DR HSP; P01180; INPO.

DR InterPro: IPR001832; Claudin.
 DR InterPro: IPR001673; S_mold_repeat.
 DR Prodom: PD006869; S_mold_repeat; 2.
 DR PROSITE: PS01346; CLAUDIN; UNKNOWN.1.
 SQ SEQUENCE 1245 AA; 135824 MW; 3249C749AFA0CDF8 CRC64;

Query Match 60.5%; Score 52; DB 3; Length 1245;
 Best Local Similarity 60.0%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 RWPMPWRK 12
 Db 1185 RQWMSWPRR 1194

RESULT 11

ID 082059 PRELIMINARY; PRT; 298 AA.
 AC 082059;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Dihydropteroate synthase.
 GN PAE2937.
 OS Pyrobaculum aerophilum.
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 OC NCBI_TaxID=13773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
 RT aerophilum."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
 DR EMBL: AE009902; AAL64549.1; -
 DR InterPro: IPR000489; Dhdropt_synth.
 DR Pfam: PF00809; Pterin_bind; 1.
 KW Complete proteome.
 SQ SEQUENCE 298 AA; 32885 MW; 0A463F36739D3ED1 CRC64;

Query Match 59.3%; Score 51; DB 17; Length 298;
 Best Local Similarity 71.4%; Pred. No. 23;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 RWPMPWRK 9
 Db 209 QWPWKKW 215

RESULT 12

ID 09Y4N1 PRELIMINARY; PRT; 299 AA.
 AC 09Y4N1;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE Hypothetical 34.0 kDa protein (Fragment).
 GN DKFP434C192.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Ansgore W., Winkler U., Mewes H.W., Gassenhuber J., Wiemann S.;
 RT Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL096753; CAB46428.2; -
 KW Hypothetical protein.
 FT NON_TER 1 1

SQ SEQUENCE 299 AA; 34032 MW; 6B8DB60E6A88239A CRC64;
 Query Match 59.3%; Score 51; DB 4; Length 299;
 Best Local Similarity 85.7%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 PMPWRR 11
 111111
 Db 37 PMPWMSR 43

RESULT 13
 P96143
 ID P96143 PRELIMINARY; PRT; 504 AA.
 AC P96143;
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
 DE Peptide hydrolase.
 GN TLEP1.
 OS Thermophilomycetes vulgaris.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Thermophilomycetaceae; Thermophilomycetes.
 ON NCBI_TaxID=2026;
 RN [1]
 RP SEQUENCE OF 1-431 FROM N.A.
 RC STRAIN=94-2A;
 RA Hofmeister J.W.;
 RL Thesis (1995), Molecular genetics,
 RL Institut fuer pflanzen-genetik und kulturpflanzenforschung.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=94-2A;
 RA Hofmeister J.W.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z83214; CAB05671.1; -.
 DR HSSP; P00800; 1HYT.
 DR InterPro; IPR001570; Peptidase_M4.
 DR Pfam; PF01447; Peptidase_M4; 1.
 DR Pfam; PF02868; Peptidase_M4_C; 1.
 DR SEQUENCE 504 AA; 56653 MW; 5A7BCC05C5AD1315 CRC64;

Query Match 59.3%; Score 51; DB 2; Length 504;
 Best Local Similarity 60.0%; Pred. No. 36;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ILRPMWMPW 10
 :11111111
 Db 71 LVRTWMPMR 80

RESULT 14
 O9CZAI
 ID O9CZAI PRELIMINARY; PRT; 141 AA.
 AC O9CZAI;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE 2810031J10RIK protein.
 GN 2810031J10RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK012846; BAB28508.1; -.
 DR MGD; MGI:1919917; 2810031J10RIK.
 DR InterPro; IPR003309; Treg-SCAN.
 DR Pfam; PF02023; SCAN; 1.
 DR SMART; SM00431; LER; 1.
 DR PROSITE; PSS0804; SCAN_BOX; 1.
 DR SEQUENCE 141 AA; 15993 MW; 865C6B735BF8203D CRC64;

Query Match 58.1%; Score 50; DB 11; Length 141;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ILRPMWMPW 9
 :111111
 Db 105 VSRRPMPW 113

RESULT 15
 O9AUN3
 ID O9AUN3 PRELIMINARY; PRT; 327 AA.
 AC O9AUN3;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Hypothetical 36.7 kDa protein.
 GN OSJNB40058E19.18.
 GN OSJNB40058E19 (Rice).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 ON NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., King L., Kirchoff K.A., de la Bastide M., Preston R.R.,
 RA Nascimento L.U., Vall M.D., Baker J.P., Miller B., Connors D.M.,
 RA Kuit K.H., Rodriguez S., Santos L., Zlatavarn T., Balija V.S.,
 RA Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
 RA McCombie W.R.;
 RT "Genomic Sequence for Oryza sativa, Nipponbare strain, Chromosome X,
 RT clone OSJNB40058E19, Complete sequence.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC083945; AKI13143.1; -.
 DR SEQUENCE 327 AA; 36672 MW; 5CCA908064BD0CA CRC64;

Query Match 58.1%; Score 50; DB 10; Length 327;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 WMPWRR 11
 11111111
 Db 119 WMPWRR 124

Search completed: January 15, 2003, 18:04:05
 Job time : 32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:07:16 : Search time 28 Seconds
(without alignments)
272.277 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKKWPMWPMRRKhaeapeapimlkkwpmwprk 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp._archaea:*
2: sp._bacteria:*
3: sp._fungi:*
4: sp._human:*
5: sp._invertebrate:*
6: sp._mammal:*
7: sp._mhc:*
8: sp._organelle:*
9: sp._phage:*
10: sp._plant:*
11: sp._rodent:*
12: sp._virus:*
13: sp._vertebrate:*
14: sp._unclassified:*
15: sp._virus:*
16: sp._bacteriophage:*
17: sp._archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	70.5	29.3	1245 3	09Y7V5
2	67.5	28.0	746 12	09JH31
3	67	27.8	723 12	09DUC4
4	64	26.6	175 12	091RD8
5	63	26.1	49 12	09DT80
6	63	26.1	192 16	0928B7
7	63	26.1	748 12	09DT81
8	63	26.1	750 12	091D04
9	62	25.7	367 11	063778
10	61	25.3	92 12	08V7E2
11	61	25.3	250 17	08TWMG
12	61	25.3	284 12	094C18
13	61	25.3	734 12	08V7I1
14	61	25.3	735 12	09DUC9
15	61	25.3	985 15	098414
16	60	24.9	147 11	061427

17	60	24.9	242 17	08TYM8	08TYM8 methanopyru
18	60	24.9	381 16	09A7E1	09A7E1 caulobacter
19	60	24.9	766 12	091CY5	091CY5 tt virus. o
20	59.5	24.7	114 16	09X8C2	09X8C2 streptomyc
21	59.5	24.7	2292 12	066765	066765 encephalomy
22	59	24.5	95 10	09LON0	091n0 arabidopsis
23	59	24.5	485 16	P72844	P72844 synechocyst
24	59	24.5	739 12	092A03	092A03 tt virus. o
25	58.5	24.3	521 20	094EF3	094EF3 oryza sativ
26	58.5	24.3	1567 2	09ADN1	09ADN1 polyanthum
27	58	24.1	107 16	09XAE4	09XAE4 streptomyc
28	58	24.1	513 11	063289	063289 rattus norv
29	58	24.1	646 11	063779	063779 rattus norv
30	58	24.1	879 11	08V199	08V199 rattus norv
31	58	24.1	1300 11	P97692	P97692 rattus norv
32	57.5	23.9	117 15	09YR9	09YR9 human immun
33	57.5	23.9	341 13	090644	090644 gallus gall
34	57.5	23.9	376 3	094516	094516 schizosacch
35	57.5	23.9	444 4	09HC40	09HC40 homo sapien
36	57.5	23.9	772 4	09BXY6	09BXY6 homo sapien
37	57	23.7	252 15	09TU37	09TU37 human immun
38	57	23.7	426 12	092A07	092A07 tt virus. o
39	57	23.7	970 11	088821	088821 mus musculi
40	57	23.7	971 11	070458	070458 mus musculi
41	56.5	23.4	157 5	09YOE8	09YOE8 drosophila
42	56.5	23.4	162 5	09W1W7	09W1W7 drosophila
43	56.5	23.4	532 16	09CK19	09CK19 pasteurella
44	56	23.2	252 15	09ITX8	09ITX8 human immun
45	56	23.2	252 15	09ITV2	09ITV2 human immun

ALIGNMENTS

RESULT 1	ID	Q9Y7V5	PRELIMINARY:	PRT: 1245 AA.
AC	Q9Y7V5:			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Conidiospore surface protein.			
GN	CMPI.			
OS	Trichoderma harzianum.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreales; mitosporic Hypocreales; Trichoderma.			
OX	NCBI_TaxID=5544;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SRPAT-ATCC 32173;			
RX	MEDLINE-99343881; PubMed=10413618;			
RA	Puyesky M., Benhamou N., Ponce Noyola P., Bauw G., Ziv T.,			
RA	Van Montagu M., Herrera Estrella A., Horwitz B.A.;			
RT	"Developmental regulation of cmpl, a gene encoding a multidomain			
RT	Conidiospore surface protein of Trichoderma.";			
RL	Fungal Genet. Biol. 27:88-99(1999).			
DR	EMBL: AJ133651; CAB40845.1; .			
DR	HSSP: P01180; INPO.			
DR	InterPro: IPR001832; Claudin.			
DR	InterPro: IPR001673; S_mold_repeat.			
DR	Prodom: PD006869; S_mold_repeat; 2.			
DR	PROSITE: PS01346; CLAUDIN; UNKNOWN_1.			
SO	SEQUENCE 1245 AA; 135824 MW; 3249C749AFA0CDF8 CRC64;			
Query Match	29.3%; Score 70.5; DB 3; Length 1245;			
Best Local Similarity	29.0%; Pred. No. 2.3;			
Matches	9; Conservative	2; Mismatches	9; Indels	11; Gaps
OY	4 KPWPMWPMRRKhaeapeapimlkkwpmwprk 34			
DB	1185 RMOQWMPWRKRG-----CMQWMSW 1204			

```
RESULT 2
09JH31      PRELIMINARY;      PRT;      746 AA.
ID 09JH31;
AC 09JH31;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUN02;
RA Okamoto H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JUN02;
RX MEDLINE=20456801; PubMed=11003468;
RA Ukita M., Okamoto H., Nishizawa T., Tawara A., Takahashi M.,
RA Iizuka H., Miyakawa Y., Mayumi M.;
RT "The entire nucleotide sequences of two distinct TT virus (TTV)
RT isolates (JUN01 and JUN02) remotely related to the original TTV
RT isolates."
RL Arch. Virol. 145:1543-1559(2000).
DR EMBL; AB028669; BAA94878.1; -.
DR InterPro: IPR004219; TTVirus_unk.
DR Pfam: PF02956; TT_ORF1.1.
SQ SEQUENCE 746 AA; 88561 MW; E0B32953AE764B3E CRC64;

Query Match      28.0%; Score 67.5; DB 12; Length 746;
Best Local Similarity 33.3%; Pred. No. 3.2;
Matches 11; Conservative 2; Mismatches 5; Indels 15; Gaps 1;

QY 5 WPMWPMRRKHAEPEAPIMILKKWPMWPMRRK 37
ID 11111111111111111111111111111111
Db 3 WGMWRMR-----RWPARRRRR 20

RESULT 3
09DUC4      PRELIMINARY;      PRT;      723 AA.
ID 09DUC4;
AC 09DUC4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RA Okamoto H.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MF-TTV9;
RX MEDLINE=20534983; PubMed=11080484;
RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,
RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
RT "Species-specific TT viruses in humans and nonhuman primates and their
RT phylogenetic relatedness."
RL Virology 277:368-378(2000).
DR EMBL; AB041959; BAB19313.1; -.
DR InterPro: IPR001563; Serine_carpept.
DR InterPro: IPR004219; TTVirus_unk.
DR Pfam: PF02956; TT_ORF1.1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
SQ SEQUENCE 723 AA; 85393 MW; 23DD003098766344 CRC64;

Query Match      27.8%; Score 67; DB 12; Length 723;
```

```
Best Local Similarity 34.4%; Pred. No. 3.6;
Matches 11; Conservative 1; Mismatches 2; Indels 18; Gaps 1;

QY 6 PMWPMRRKHAEPEAPIMILKKWPMWPMRRK 37
ID 11111111111111111111111111111111
Db 2 PMWPMRR-----WRRWRRR 15

RESULT 4
091RD8      PRELIMINARY;      PRT;      175 AA.
ID 091RD8;
AC 091RD8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ORF3.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L03;
RA Liu Z.H., Luo K.X., Hu J., He H.T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF371370; AAK54733.1; -.
DR InterPro: IPR004219; TTVirus_unk.
DR Pfam: PF02956; TT_ORF1.1.
SQ SEQUENCE 175 AA; 22073 MW; 5212D7DA3FD72F81 CRC64;

Query Match      26.6%; Score 64; DB 12; Length 175;
Best Local Similarity 33.3%; Pred. No. 2.1;
Matches 11; Conservative 2; Mismatches 4; Indels 16; Gaps 2;

QY 5 WPMWPMRRKHAEPEAPIMILKKWPMWPMRRK 37
ID 11111111111111111111111111111111
Db 3 WSWW-WRRRR-----WMPRRRR 19

RESULT 5
09DT80      PRELIMINARY;      PRT;      49 AA.
ID 09DT80;
AC 09DT80;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ORF1 (Fragment).
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYM9;
RX MEDLINE=20568739; PubMed=11118348;
RA Okamoto H., Nishizawa T., Tawara A., Takahashi M., Kishimoto J.,
RA Sai T., Sugai Y.;
RT "TT virus mRNAs detected in the bone marrow cells from an infected
RT individual."
RL Biochem. Biophys. Res. Commun. 279:700-707(2000).
DR EMBL; AB050449; BAB19930.1; -.
DR NON_TER 49
SQ SEQUENCE 49 AA; 7225 MW; 1DA6F8F1AB69AA43 CRC64;

Query Match      26.1%; Score 63; DB 12; Length 49;
Best Local Similarity 30.3%; Pred. No. 0.79;
Matches 10; Conservative 3; Mismatches 4; Indels 16; Gaps 2;

QY 5 WPMWPMRRKHAEPEAPIMILKKWPMWPMRRK 37
ID 11111111111111111111111111111111
Db 3 WTW-WQRRRR-----WPMRRR 19

RESULT 6
09ZBB7
```


SEQUENCE / 34 AA; 809 / 8 MW; F00E180BC01047A00 CINC07

query	Match	Score	DB	Length
25.38;	Score 61;	DB 12;	Length 734;	

Best Local Similarity 32.3%; Pred. No. 19;
Matches 10; Conservative 3; Mismatches 2; Indels 16; Gaps 1;

QY 7 WPMRRKHEAPEAPIMILKKMPWPRRK 37
DB 3 WMYRRR-----PWRPWRRR 17

RESULT 14

Q9DUC9 PRELIMINARY; PRT; 735 AA.

AC Q9DUC9;

DT 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE ORF1.

OS TT virus.

OC Viruses; ssDNA viruses; unclassified ssDNA viruses.

OX NCBI_Taxid=68887;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PT-TTV6;

RA Okamoto H.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=PT-TTV6;

RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,

RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;

RT "Species-specific TT viruses in humans and nonhuman primates and their

RT phylogenetic relatedness.";

RL Virology 277:368-378(2000).

DR EMBL: AB041957; BAB19308.1;

DR InterPro: IPR004219; TTVirus_Unk.

DR Pfam: PF02956; TT_ORF1; 1.

SQ SEQUENCE 735 AA; 86132 MW; 9ED818D6B6E6FA5D3 CRC64;

Query Match 25.3%; Score 61; DB 12; Length 735;

Best Local Similarity 29.3%; Pred. No. 19;

Matches 12; Conservative 3; Mismatches 6; Indels 20; Gaps 3;

QY 5 WPMRRKHEAPEAPIMILKKMPWPRRK 37
DB 3 WPMRRRKRWRRRR-----PWRRRRRRTWRRR 31

RESULT 15

Q98414 PRELIMINARY; PRT; 985 AA.

AC Q98414;

DT 01-FEB-1997 (TREMBlrel. 02, Created)

DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Envelope glycoprotein.

GN ENV.

OS Ovine lentivirus.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_Taxid=11663;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=85/34;

RA MEDLINE=95135990; Pubmed=7834396;

RA Woodward T.M., Carlson J.O., de la Concha-Bermejo A.,

RA Demartini J.C.;

RT "Biological and genetic changes in ovine lentivirus strains following

RT passage in isogenic twin lambs.";

RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 8:124-133(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=85/34;

RA Carlson J.O., Demartini J.C., Mwaengo D.M.;

RT "Envelope glycoprotein nucleotide sequence and genetic

RT Characterization of North American ovine lentiviruses.";

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U64439; AB08725.1;

DR InterPro: IPR000328; Env_GP41.

DR Pfam: PF00517; GP41; 1.

KW Transmembrane.

SQ SEQUENCE 985 AA; 113794 MW; 3197258BDBE3597 CRC64;

Query Match 25.3%; Score 61; DB 15; Length 985;

Best Local Similarity 23.4%; Pred. No. 25;

Matches 18; Conservative 7; Mismatches 8; Indels 44; Gaps 5;

QY 1 ILKK-----WP-----WRRKHEAPEAPE----- 21
DB 163 ILKRVYKODWPWNTYHWPWMENNRQWKEEREYKGRNKTEDIDDLAKIRGRC 222

QY 22 --PIMLK--KWPWP 33
DB 223 VPYPFALLKCTKWCWYP 239

Search completed: January 15, 2003, 18:09:29
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:04:50 : Search time 11 Seconds

(without alignments)
139.511 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKKWPMWPMRRKheapeapeimlkkkmpwprk 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	78	32.4	144	INDC_BOVIN
2	62	25.7	314	YMA3_BACST
3	59.5	24.7	2290	POLG_EMCV
4	57	23.7	711	MMIA_SPRO
5	55.5	23.0	55	ATP8_ANAPL
6	55.5	23.0	942	ENY_CAEVG
7	55.5	23.0	990	ENY_OMVVS
8	55	22.8	715	YD55_MYCTU
9	54	22.4	144	AP22_APIME
10	54	22.4	469	SYCL_MYCTU
11	54	22.4	473	SYCL_MYCTU
12	54	22.4	1173	VGL2_CVH22
13	53.5	22.2	1042	CORI_HOMAN
14	53.5	22.2	1113	CORI_MOUSE
15	53	22.0	424	FDP6_SOYBN
16	52.5	21.8	55	ATP8_AYTAM
17	52.5	21.8	257	E434_ADE40
18	52.5	21.8	691	YH3G_ECOLI
19	52.5	21.8	982	ENY_VILV
20	52.5	21.8	983	ENY_VILV
21	52.5	21.8	991	ENY_VILV
22	52.5	21.8	1154	VGL2_IBVD2
23	52.5	21.8	1162	VGL2_IBVD
24	52.5	21.8	1162	VGL2_IBV
25	52.5	21.8	1162	VGL2_IBV
26	52.5	21.8	1162	VGL2_IBV
27	52	21.6	68	Y121_BPT4
28	51.5	21.4	162	DSBB_NEIMA
29	51.5	21.4	162	DSBB_NEIMA
30	51.5	21.4	175	MUCG_ANASP
31	51.5	21.4	443	FDP6_BRANA
32	51.5	21.4	448	FDP6_ARATH
33	51.5	21.4	989	ENY_VILV1

ALIGNMENTS

RESULT 1	ID	INDC_BOVIN	STANDARD:	PRT:	144 AA.
AC	P33046;				
DT	01-OCT-1993 (Rel. 27, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Indolicidin precursor.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone marrow;				
RA	Medline-92392368; PubMed-1520337;				
RA	del Sal G., Storici P., Schneider C., Romeo D., Zanetti M.;				
RT	"CDNA cloning of the neutrophil bactericidal peptide indolicidin."				
RL	Biochem. Biophys. Res. Commun. 187:467-472(1992).				
RN	[2]				
RP	SEQUENCE OF 131-143.				
RC	TISSUE=Neutrophils;				
RA	Medline-92165771; PubMed-1537821;				
RA	Selsted M.E., Novotny M.J., Morris W.L., Tang Y.-Q., Smith W.;				
RT	Cullor J.S.;				
RT	"Indolicidin, a novel bactericidal tridecapeptide amide from				
RT	neutrophils."				
RL	J. Biol. Chem. 267:4292-4295(1992).				
CC	- FUNCTION: POTENT MICROBICIDAL ACTIVITY, ACTIVE AGAINST				
CC	STAPHYLOCOCCUS AUREUS AND ESCHERICHIA COLI.				
CC	- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.				
CC	- PFM: ELASTASE MIGHT BE RESPONSIBLE FOR ITS MATURATION.				
CC	- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: X67340; CAA47755.1; -				
CC	PIR: JCI1222; JCI1222.				
CC	PIR: A42387; A42387.				
DR	InterPro: IPR001894; Cathelicidin.				
DR	Pfam: PF00666; Cathelicidins; 1.				
DR	ProDom: PD001838; Cathelicidin; 1.				
DR	PROSITE: PS00946; CATHELICIDINS_1; 1.				
DR	PROSITE: PS00947; CATHELICIDINS_2; 1.				
KW	Antibiotic; Amidation; Signal				
FT	SIGNAL 1 29				POTENTIAL.
FT	PROPEP 30 130				INDOLICIDIN.
FT	PEPTIDE 131 143				PYRROLIDONE CARBOXYLIC ACID (BY
FT	MOD_RES 30 30				

```

FT DISULFID 85 96 BY SIMILARITY.
FT DISULFID 107 124 BY SIMILARITY.
FT MOD_RES 143 143 AMIDATION (G-144 PROVIDE AMIDE GROUP).
SQ SEQUENCE 144 AA; 16479 MW; E3B1CBHE5C09911 CRC64;

Query Match
Best Local Similarity 32.4%; Score 78; DB 1; Length 144;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 19 EAEPIMILKKPMPWMPWR 36
+ : : : : :
Db 126 ELQSVLLPKKMPWMPWR 143

RESULT 2
YMA3_BACST
ID YMA3_BACST STANDARD; PRT; 314 AA.
AC 045633;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 35.7 kDa protein in mala 3' region (ORF3).
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 7953;
RC MEDLINE=94247374; PubMed=8190087;
RA Liang E.C., Ferencik T.;
RT "Molecular cloning of a maltose transport gene from Bacillus
steatothermophilus and its expression in Escherichia coli K-12.";
RL Mol. Gen. Genet. 243:343-352(1994).
CC -1- SIMILARITY: BELONGS TO THE UPF0097 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC
CC
CC EMBL: L13418; AAA71981.1; -
DR InterPro: IPR004175; 2_5_1ligase.
DR Pfam: PF02834; 2_5_1ligase; 2.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 35735 MW; B54E25FD3F72BFC4 CRC64;

Query Match
Best Local Similarity 25.7%; Score 62; DB 1; Length 314;
Matches 17; Conservative 5; Mismatches 13; Indels 14; Gaps 3;

OY 2 LKKPMPWPKRRKHEAPEA-----EP-----IMLKKMPWMP-----WPR 36
+ : : : : :
Db 222 LKQWQDAKARWEARDADGAREPGFTWDFLLDKPSWMERLIGRWRR 270

RESULT 3
POLG_EMCV
ID POLG_EMCV STANDARD; PRT; 2290 AA.
AC P03304;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins
P2A TO P2C; P3A; Genome-linked protein VPg; Picornain 3C
DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
DE (EC 2.7.7.48)].
OS Enterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
Cardiovirus.

```

```

OX NCBI_TaxID=12104;
RN (1)
RP SEQUENCE FROM N.A.
RC MEDLINE=84169586; PubMed=6324136;
RC Patemburg A.C., Kirby E.M., Janda M.R., Drake N.L., Duke G.M.,
RA Potratz K.F., Collett M.S.;
RT "The nucleotide and deduced amino acid sequences of the
encephalomyocarditis viral polypeptide coding region.";
RL Nucleic Acids Res. 12:2969-2985(1984).
CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the
poliovirus polypeptide. In other picornavirus reactions Gln may be
substituted for Glu, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
[RNA](N).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: X00463; CAA25152.1; -
DR PIR: A03906; GNNE.
DR HSSP: P12296; 2MEV.
DR MEROPS: C03.009; -
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00073; RV; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polypeptide; Coat protein; Core protein; Transferase;
RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
KW PROPEP
FT CHAIN 1 67 LEADER PEPTIDE.
FT CHAIN 68 136 COAT PROTEIN VP4 (RH0).
FT CHAIN 137 391 COAT PROTEIN VP2 (BETA).
FT CHAIN 392 622 COAT PROTEIN VP3 (GAMMA).
FT CHAIN 623 910 COAT PROTEIN VP1 (ALPHA).
FT CHAIN 911 1056 COAT PROTEIN P2A (G).
FT CHAIN 1057 1192 CORE PROTEIN P2B (F).
FT CHAIN 1193 1517 CORE PROTEIN P2C (F).
FT CHAIN 1518 1605 CORE PROTEIN P3A.
FT CHAIN 1606 1625 GENOME-LINKED PROTEIN VPg (H).
FT CHAIN 1626 1830 PICORNAIN 3C (P22).
FT CHAIN 1831 2290 RNA-DIRECTED RNA POLYMERASE P3D (E).
FT LIPID 68 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1784 1784 PROTEASE (POTENTIAL).
FT ACT_SITE 1802 1802 PROTEASE (POTENTIAL).
SQ SEQUENCE 2290 AA; 255756 MW; 26BC81BB7CE68C5 CRC64;

Query Match
Best Local Similarity 24.7%; Score 59.5; DB 1; Length 2290;
Matches 11; Conservative 6; Mismatches 8; Indels 17; Gaps 1;

OY 6 PMPWPKRRKHEAPEAEP-----MILKKMP 30
+ : : : : :
Db 967 PMPWPKRRKHEAPEAEP-----MILKKMP 1008

RESULT 4
MMLA_STRCO
ID MMLA_STRCO STANDARD; PRT; 711 AA.
AC Q53902;
DT 30-MAY-2000 (Rel. 39, Created)

```


QY	1	ILKKPMMWPKRKHEAPE	19
DB	332	IFGRWVMPARKHGTETPD	350

30-MAY-2000 (Rel. 39, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)

Putative membrane protein act11-3.

ACT11-3 OR SCO5084 OR SCBAC2861.10.

Streptomyces coelicolor.

Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

NCBI_TaxID=1902;

[1]

SEQUENCE FROM N.A.

RA MEDLINE=91347376; PubMed=1878971;

RA Fernandez-Moreno M.A., Caballero J.L., Hopwood D.A., Malpartida F.,

RT "The act cluster contains regulatory and antibiotic export genes,

RT direct targets for translational control by the bida tRNA gene of

Streptomyces";

Cell 66:769-780(1991).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdeno-Parraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Raddiolowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RL coelicolor A3(2)";

Nature 417:141-147(2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE MPMPL FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; M64683; AAA26691.1; -;

DR EMBL; AL593842; CAC44197.1; -;

DR InterPro: IPR004707; Act11.

DR InterPro: IPR000731; HMGCR/patch_5TM.

DR InterPro: IPR004869; MPMPL.

DR Pfam: PF031176; MPMPL_2.

DR TIGRFAMs: TIGRF00833; act11.1.

DR PROSITE; PS50156; SSD; 2.

DR Hypothetical protein; Transmembrane; Complete proteome.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 14 34 POTENTIAL.

FT TRANSMEM 175 195 POTENTIAL.

FT TRANSMEM 199 219 POTENTIAL.

FT TRANSMEM 235 255 POTENTIAL.

FT TRANSMEM 281 301 POTENTIAL.

FT TRANSMEM 313 333 POTENTIAL.

FT TRANSMEM 369 389 POTENTIAL.

FT TRANSMEM 516 536 POTENTIAL.

FT TRANSMEM 540 560 POTENTIAL.

FT TRANSMEM 573 593 POTENTIAL.

FT TRANSMEM 623 643 POTENTIAL.

FT TRANSMEM 645 665 POTENTIAL.

SO SEQUENCE 711 AA; 74862 MW; A5466E0DA0BEDD1B6 CAC64;

Query Match 23.7%; Score 57; DB 1; Length 711;

Best Local Similarity 47.4%; Pred. No. 9.1;

Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```

RESULT 5
ATPB_AMPPL
ID      ATPB_AMPPL      STANDARD:      PRT:      55 AA.
AC      P50655.
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6U).
GN      MYPAP8 OR ATPB8.
OS      Anas platyrhynchos (Domestic duck).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauromia; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX      NCBI_TaxID=8839.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=pekin breed, TISSUE=Liver;
RX      MEDLINE=94047124; PubMed=8230253;
RX      Ramirez V., Savole P., Morais R.;
RT      "Molecular characterization and evolution of a duck mitochondrial
RT      genome.";
RL      J. Mol. Evol. 37:296-310(1993).
CC      -I- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC      (CE(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC      -I- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC      H(+) (out).
CC      -I- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: L22476; AAA72037.1; -
DR      InterPro: IPR001421; ATPase8_mlt.
DR      Pfam: PF00895; ATP-synt_8; 1.
KW      Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT      TRANSMEM      8      24
SQ      SEQUENCE      55 AA: 6334 MW: 41284ABAB1525599 CRC64;
      -----
Query Match      23.0%, Score 55.5; DB 1; Length 55;
Best local Similarity 50.0%; Pred. No. 1.1;
Matches 9; Conservative 2; Mismatches 6; Indels 1; Gaps 1;
Oy      18 PEAEPIMLTKKPPW-WPW 34
      | : | | | | | |
Db      37 PSKKPSLITTKPPWAMPW 54

RESULT 6
ENV_CAEVG
ID      ENV_CAEVG      STANDARD:      PRT:      942 AA.
AC      P31627.
DT      01-JUL-1993 (Rel. 26, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Env polyprotein precursor (Coat polyprotein) [Contains: Surface
DE      protein; Transmembrane protein].
GN      ENV.
OS      Caprine arthritis encephalitis virus (strain 663) (CAEV).
OC      Viruses; Retroid viruses; Retroviridae; Lentiviruses.
OC      NCBI_TaxID=11662.
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 81-95.
RX      MEDLINE=92015464; PubMed=1656067.
RX      Knowles D.P., Jr., Cheevers W.P., McGuire T.C., Brassfield A.L.,
RX      Harwood W.G., Stem T.A.;

```

RT "Structure and genetic variability of envelope glycoproteins of two
RT antigenic variants of caprine arthritis-encephalitis lentivirus."
RL J. Virol. 65:5744-5750(1991).
[2]
RP REVISIONS.
RA Knowles D.P.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M60855; AAB88709.2; -
DR PIR: A41307; VGLJG6.
DR InterPro: IPR000328; Env_GPA1.
DR Pfam: PF00517; GPA1; 1.
KW Glycoprotein; Coat protein; Polypeptide; Transmembrane.
FT PEPTIDE 1 80
FT CHAIN 81 630
FT DOMAIN 1 630
FT TRANSMEM 631 659
FT DOMAIN 660 799
FT TRANSMEM 800 820
FT DOMAIN 821 942
FT CARBOHYD 51 98
FT CARBOHYD 98 131
FT CARBOHYD 131 176
FT CARBOHYD 176 228
FT CARBOHYD 228 331
FT CARBOHYD 331 348
FT CARBOHYD 348 354
FT CARBOHYD 354 370
FT CARBOHYD 370 379
FT CARBOHYD 379 400
FT CARBOHYD 400 404
FT CARBOHYD 404 435
FT CARBOHYD 435 441
FT CARBOHYD 441 447
FT CARBOHYD 447 457
FT CARBOHYD 457 467
FT CARBOHYD 467 481
FT CARBOHYD 481 493
FT CARBOHYD 493 503
FT CARBOHYD 503 509
FT CARBOHYD 509 527
FT CARBOHYD 527 534
FT CARBOHYD 534 544
FT CARBOHYD 544 554
FT CARBOHYD 554 560
FT CARBOHYD 560 567
FT CARBOHYD 567 573
FT CARBOHYD 573 577
FT CARBOHYD 577 581
FT CARBOHYD 581 585
FT CARBOHYD 585 589
FT CARBOHYD 589 593
FT CARBOHYD 593 597
FT CARBOHYD 597 601
FT CARBOHYD 601 605
FT CARBOHYD 605 609
FT CARBOHYD 609 613
FT CARBOHYD 613 617
FT CARBOHYD 617 621
FT CARBOHYD 621 625
FT CARBOHYD 625 629
FT CARBOHYD 629 633
FT CARBOHYD 633 637
FT CARBOHYD 637 641
FT CARBOHYD 641 645
FT CARBOHYD 645 649
FT CARBOHYD 649 653
FT CARBOHYD 653 657
FT CARBOHYD 657 661
FT CARBOHYD 661 665
FT CARBOHYD 665 669
FT CARBOHYD 669 673
FT CARBOHYD 673 677
FT CARBOHYD 677 681
FT CARBOHYD 681 685
FT CARBOHYD 685 689
FT CARBOHYD 689 693
FT CARBOHYD 693 697
FT CARBOHYD 697 701
FT CARBOHYD 701 705
FT CARBOHYD 705 709
FT CARBOHYD 709 713
FT CARBOHYD 713 717
FT CARBOHYD 717 721
FT CARBOHYD 721 725
FT CARBOHYD 725 729
FT CARBOHYD 729 733
FT CARBOHYD 733 737
FT CARBOHYD 737 741
FT CARBOHYD 741 745
FT CARBOHYD 745 749
FT CARBOHYD 749 753
FT CARBOHYD 753 757
FT CARBOHYD 757 761
FT CARBOHYD 761 765
FT CARBOHYD 765 769
FT CARBOHYD 769 773
FT CARBOHYD 773 777
FT CARBOHYD 777 781
FT CARBOHYD 781 785
FT CARBOHYD 785 789
FT CARBOHYD 789 793
FT CARBOHYD 793 797
FT CARBOHYD 797 801
FT CARBOHYD 801 805
FT CARBOHYD 805 809
FT CARBOHYD 809 813
FT CARBOHYD 813 817
FT CARBOHYD 817 821
FT CARBOHYD 821 825
FT CARBOHYD 825 829
FT CARBOHYD 829 833
FT CARBOHYD 833 837
FT CARBOHYD 837 841
FT CARBOHYD 841 845
FT CARBOHYD 845 849
FT CARBOHYD 849 853
FT CARBOHYD 853 857
FT CARBOHYD 857 861
FT CARBOHYD 861 865
FT CARBOHYD 865 869
FT CARBOHYD 869 873
FT CARBOHYD 873 877
FT CARBOHYD 877 881
FT CARBOHYD 881 885
FT CARBOHYD 885 889
FT CARBOHYD 889 893
FT CARBOHYD 893 897
FT CARBOHYD 897 901
FT CARBOHYD 901 905
FT CARBOHYD 905 909
FT CARBOHYD 909 913
FT CARBOHYD 913 917
FT CARBOHYD 917 921
FT CARBOHYD 921 925
FT CARBOHYD 925 929
FT CARBOHYD 929 933
FT CARBOHYD 933 937
FT CARBOHYD 937 941
FT CARBOHYD 941 945
FT CARBOHYD 945 949
FT CARBOHYD 949 953
FT CARBOHYD 953 957
FT CARBOHYD 957 961
FT CARBOHYD 961 965
FT CARBOHYD 965 969
FT CARBOHYD 969 973
FT CARBOHYD 973 977
FT CARBOHYD 977 981
FT CARBOHYD 981 985
FT CARBOHYD 985 989
FT CARBOHYD 989 993
FT CARBOHYD 993 997
FT CARBOHYD 997 1001
FT CARBOHYD 1001 1005
FT CARBOHYD 1005 1009
FT CARBOHYD 1009 1013
FT CARBOHYD 1013 1017
FT CARBOHYD 1017 1021
FT CARBOHYD 1021 1025
FT CARBOHYD 1025 1029
FT CARBOHYD 1029 1033
FT CARBOHYD 1033 1037
FT CARBOHYD 1037 1041
FT CARBOHYD 1041 1045
FT CARBOHYD 1045 1049
FT CARBOHYD 1049 1053
FT CARBOHYD 1053 1057
FT CARBOHYD 1057 1061
FT CARBOHYD 1061 1065
FT CARBOHYD 1065 1069
FT CARBOHYD 1069 1073
FT CARBOHYD 1073 1077
FT CARBOHYD 1077 1081
FT CARBOHYD 1081 1085
FT CARBOHYD 1085 1089
FT CARBOHYD 1089 1093
FT CARBOHYD 1093 1097
FT CARBOHYD 1097 1101
FT CARBOHYD 1101 1105
FT CARBOHYD 1105 1109
FT CARBOHYD 1109 1113
FT CARBOHYD 1113 1117
FT CARBOHYD 1117 1121
FT CARBOHYD 1121 1125
FT CARBOHYD 1125 1129
FT CARBOHYD 1129 1133
FT CARBOHYD 1133 1137
FT CARBOHYD 1137 1141
FT CARBOHYD 1141 1145
FT CARBOHYD 1145 1149
FT CARBOHYD 1149 1153
FT CARBOHYD 1153 1157
FT CARBOHYD 1157 1161
FT CARBOHYD 1161 1165
FT CARBOHYD 1165 1169
FT CARBOHYD 1169 1173
FT CARBOHYD 1173 1177
FT CARBOHYD 1177 1181
FT CARBOHYD 1181 1185
FT CARBOHYD 1185 1189
FT CARBOHYD 1189 1193
FT CARBOHYD 1193 1197
FT CARBOHYD 1197 1201
FT CARBOHYD 1201 1205
FT CARBOHYD 1205 1209
FT CARBOHYD 1209 1213
FT CARBOHYD 1213 1217
FT CARBOHYD 1217 1221
FT CARBOHYD 1221 1225
FT CARBOHYD 1225 1229
FT CARBOHYD 1229 1233
FT CARBOHYD 1233 1237
FT CARBOHYD 1237 1241
FT CARBOHYD 1241 1245
FT CARBOHYD 1245 1249
FT CARBOHYD 1249 1253
FT CARBOHYD 1253 1257
FT CARBOHYD 1257 1261
FT CARBOHYD 1261 1265
FT CARBOHYD 1265 1269
FT CARBOHYD 1269 1273
FT CARBOHYD 1273 1277
FT CARBOHYD 1277 1281
FT CARBOHYD 1281 1285
FT CARBOHYD 1285 1289
FT CARBOHYD 1289 1293
FT CARBOHYD 1293 1297
FT CARBOHYD 1297 1301
FT CARBOHYD 1301 1305
FT CARBOHYD 1305 1309
FT CARBOHYD 1309 1313
FT CARBOHYD 1313 1317
FT CARBOHYD 1317 1321
FT CARBOHYD 1321 1325
FT CARBOHYD 1325 1329
FT CARBOHYD 1329 1333
FT CARBOHYD 1333 1337
FT CARBOHYD 1337 1341
FT CARBOHYD 1341 1345
FT CARBOHYD 1345 1349
FT CARBOHYD 1349 1353
FT CARBOHYD 1353 1357
FT CARBOHYD 1357 1361
FT CARBOHYD 1361 1365
FT CARBOHYD 1365 1369
FT CARBOHYD 1369 1373
FT CARBOHYD 1373 1377
FT CARBOHYD 1377 1381
FT CARBOHYD 1381 1385
FT CARBOHYD 1385 1389
FT CARBOHYD 1389 1393
FT CARBOHYD 1393 1397
FT CARBOHYD 1397 1401
FT CARBOHYD 1401 1405
FT CARBOHYD 1405 1409
FT CARBOHYD 1409 1413
FT CARBOHYD 1413 1417
FT CARBOHYD 1417 1421
FT CARBOHYD 1421 1425
FT CARBOHYD 1425 1429
FT CARBOHYD 1429 1433
FT CARBOHYD 1433 1437
FT CARBOHYD 1437 1441
FT CARBOHYD 1441 1445
FT CARBOHYD 1445 1449
FT CARBOHYD 1449 1453
FT CARBOHYD 1453 1457
FT CARBOHYD 1457 1461
FT CARBOHYD 1461 1465
FT CARBOHYD 1465 1469
FT CARBOHYD 1469 1473
FT CARBOHYD 1473 1477
FT CARBOHYD 1477 1481
FT CARBOHYD 1481 1485
FT CARBOHYD 1485 1489
FT CARBOHYD 1489 1493
FT CARBOHYD 1493 1497
FT CARBOHYD 1497 1501
FT CARBOHYD 1501 1505
FT CARBOHYD 1505 1509
FT CARBOHYD 1509 1513
FT CARBOHYD 1513 1517
FT CARBOHYD 1517 1521
FT CARBOHYD 1521 1525
FT CARBOHYD 1525 1529
FT CARBOHYD 1529 1533
FT CARBOHYD 1533 1537
FT CARBOHYD 1537 1541
FT CARBOHYD 1541 1545
FT CARBOHYD 1545 1549
FT CARBOHYD 1549 1553
FT CARBOHYD 1553 1557
FT CARBOHYD 1557 1561
FT CARBOHYD 1561 1565
FT CARBOHYD 1565 1569
FT CARBOHYD 1569 1573
FT CARBOHYD 1573 1577
FT CARBOHYD 1577 1581
FT CARBOHYD 1581 1585
FT CARBOHYD 1585 1589
FT CARBOHYD 1589 1593
FT CARBOHYD 1593 1597
FT CARBOHYD 1597 1601
FT CARBOHYD 1601 1605
FT CARBOHYD 1605 1609
FT CARBOHYD 1609 1613
FT CARBOHYD 1613 1617
FT CARBOHYD 1617 1621
FT CARBOHYD 1621 1625
FT CARBOHYD 1625 1629
FT CARBOHYD 1629 1633
FT CARBOHYD 1633 1637
FT CARBOHYD 1637 1641
FT CARBOHYD 1641 1645
FT CARBOHYD 1645 1649
FT CARBOHYD 1649 1653
FT CARBOHYD 1653 1657
FT CARBOHYD 1657 1661
FT CARBOHYD 1661 1665
FT CARBOHYD 1665 1669
FT CARBOHYD 1669 1673
FT CARBOHYD 1673 1677
FT CARBOHYD 1677 1681
FT CARBOHYD 1681 1685
FT CARBOHYD 1685 1689
FT CARBOHYD 1689 1693
FT CARBOHYD 1693 1697
FT CARBOHYD 1697 1701
FT CARBOHYD 1701 1705
FT CARBOHYD 1705 1709
FT CARBOHYD 1709 1713
FT CARBOHYD 1713 1717
FT CARBOHYD 1717 1721
FT CARBOHYD 1721 1725
FT CARBOHYD 1725 1729
FT CARBOHYD 1729 1733
FT CARBOHYD 1733 1737
FT CARBOHYD 1737 1741
FT CARBOHYD 1741 1745
FT CARBOHYD 1745 1749
FT CARBOHYD 1749 1753
FT CARBOHYD 1753 1757
FT CARBOHYD 1757 1761
FT CARBOHYD 1761 1765
FT CARBOHYD 1765 1769
FT CARBOHYD 1769 1773
FT CARBOHYD 1773 1777
FT CARBOHYD 1777 1781
FT CARBOHYD 1781 1785
FT CARBOHYD 1785 1789
FT CARBOHYD 1789 1793
FT CARBOHYD 1793 1797
FT CARBOHYD 1797 1801
FT CARBOHYD 1801 1805
FT CARBOHYD 1805 1809
FT CARBOHYD 1809 1813
FT CARBOHYD 1813 1817
FT CARBOHYD 1817 1821
FT CARBOHYD 1821 1825
FT CARBOHYD 1825 1829
FT CARBOHYD 1829 1833
FT CARBOHYD 1833 1837
FT CARBOHYD 1837 1841
FT CARBOHYD 1841 1845
FT CARBOHYD 1845 1849
FT CARBOHYD 1849 1853
FT CARBOHYD 1853 1857
FT CARBOHYD 1857 1861
FT CARBOHYD 1861 1865
FT CARBOHYD 1865 1869
FT CARBOHYD 1869 1873
FT CARBOHYD 1873 1877
FT CARBOHYD 1877 1881
FT CARBOHYD 1881 1885
FT CARBOHYD 1885 1889
FT CARBOHYD 1889 1893
FT CARBOHYD 1893 1897
FT CARBOHYD 1897 1901
FT CARBOHYD 1901 1905
FT CARBOHYD 1905 1909
FT CARBOHYD 1909 1913
FT CARBOHYD 1913 1917
FT CARBOHYD 1917 1921
FT CARBOHYD 1921 1925
FT CARBOHYD 1925 1929
FT CARBOHYD 1929 1933
FT CARBOHYD 1933 1937
FT CARBOHYD 1937 1941
FT CARBOHYD 1941 1945
FT CARBOHYD 1945 1949
FT CARBOHYD 1949 1953
FT CARBOHYD 1953 1957
FT CARBOHYD 1957 1961
FT CARBOHYD 1961 1965
FT CARBOHYD 1965 1969
FT CARBOHYD 1969 1973
FT CARBOHYD 1973 1977
FT CARBOHYD 1977 1981
FT CARBOHYD 1981 1985
FT CARBOHYD 1985 1989
FT CARBOHYD 1989 1993
FT CARBOHYD 1993 1997
FT CARBOHYD 1997 2001
FT CARBOHYD 2001 2005
FT CARBOHYD 2005 2009
FT CARBOHYD 2009 2013
FT CARBOHYD 2013 2017
FT CARBOHYD 2017 2021
FT CARBOHYD 2021 2025
FT CARBOHYD 2025 2029
FT CARBOHYD 2029 2033
FT CARBOHYD 2033 2037
FT CARBOHYD 2037 2041
FT CARBOHYD 2041 2045
FT CARBOHYD 2045 2049
FT CARBOHYD 2049 2053
FT CARBOHYD 2053 2057
FT CARBOHYD 2057 2061
FT CARBOHYD 2061 2065
FT CARBOHYD 2065 2069
FT CARBOHYD 2069 2073
FT CARBOHYD 2073 2077
FT CARBOHYD 2077 2081
FT CARBOHYD 2081 2085
FT CARBOHYD 2085 2089
FT CARBOHYD 2089 2093
FT CARBOHYD 2093 2097
FT CARBOHYD 2097 2101
FT CARBOHYD 2101 2105
FT CARBOHYD 2105 2109
FT CARBOHYD 2109 2113
FT CARBOHYD 2113 2117
FT CARBOHYD 2117 2121
FT CARBOHYD 2121 2125
FT CARBOHYD 2125 2129
FT CARBOHYD 2129 2133
FT CARBOHYD 2133 2137
FT CARBOHYD 2137 2141
FT CARBOHYD 2141 2145
FT CARBOHYD 2145 2149
FT CARBOHYD 2149 2153
FT CARBOHYD 2153 2157
FT CARBOHYD 2157 2161
FT CARBOHYD 2161 2165
FT CARBOHYD 2165 2169
FT CARBOHYD 2169 2173
FT CARBOHYD 2173 2177
FT CARBOHYD 2177 2181
FT CARBOHYD 2181 2185
FT CARBOHYD 2185 2189
FT CARBOHYD 2189 2193
FT CARBOHYD 2193 2197
FT CARBOHYD 2197 2201
FT CARBOHYD 2201 2205
FT CARBOHYD 2205 2209
FT CARBOHYD 2209 2213
FT CARBOHYD 2213 2217
FT CARBOHYD 2217 2221
FT CARBOHYD 2221 2225
FT CARBOHYD 2225 2229
FT CARBOHYD 2229 2233
FT CARBOHYD 2233 2237
FT CARBOHYD 2237 2241
FT CARBOHYD 2241 2245
FT CARBOHYD 2245 2249
FT CARBOHYD 2249 2253
FT CARBOHYD 2253 2257
FT CARBOHYD 2257 2261
FT CARBOHYD 2261 2265
FT CARBOHYD 2265 2269
FT CARBOHYD 2269 2273
FT CARBOHYD 2273 2277
FT CARBOHYD 2277 2281
FT CARBOHYD 2281 2285
FT CARBOHYD 2285 2289
FT CARBOHYD 2289 2293
FT CARBOHYD 2293 2297
FT CARBOHYD 2297 2301
FT CARBOHYD 2301 2305
FT CARBOHYD 2305 2309
FT CARBOHYD 2309 2313
FT CARBOHYD 2313 2317
FT CARBOHYD 2317 2321
FT CARBOHYD 2321 2325
FT CARBOHYD 2325 2329
FT CARBOHYD 2329 2333
FT CARBOHYD 2333 2337
FT CARBOHYD 2337 2341
FT CARBOHYD 2341 2345
FT CARBOHYD 2345 2349
FT CARBOHYD 2349 2353
FT CARBOHYD 2353 2357
FT CARBOHYD 2357 2361
FT CARBOHYD 2361 2365
FT CARBOHYD 2365 2369
FT CARBOHYD 2369 2373
FT CARBOHYD 2373 2377
FT CARBOHYD 2377 2381
FT CARBOHYD 2381 2385
FT CARBOHYD 2385 2389
FT CARBOHYD 2389 2393
FT CARBOHYD 2393 2397
FT CARBOHYD 2397 2401
FT CARBOHYD 2401 2405
FT CARBOHYD 2405 2409
FT CARBOHYD 2409 2413
FT CARBOHYD 2413 2417
FT CARBOHYD 2417 2421
FT CARBOHYD 2421 2425
FT CARBOHYD 2425 2429
FT CARBOHYD 2429 2433
FT CARBOHYD 2433 2437
FT CARBOHYD 2437 2441
FT CARBOHYD 2441 2445
FT CARBOHYD 2445 2449
FT CARBOHYD 2449 2453
FT CARBOHYD 2453 2457
FT CARBOHYD 2457 2461
FT CARBOHYD 2461 2465
FT CARBOHYD 2465 2469
FT CARBOHYD 2469 2473
FT CARBOHYD 2473 2477
FT CARBOHYD 2477 2481
FT CARBOHYD 2481 2485
FT CARBOHYD 2485 2489
FT CARBOHYD 2489 2493
FT CARBOHYD 2493 2497
FT CARBOHYD 2497 2501
FT CARBOHYD 2501 2505
FT CARBOHYD 2505 2509
FT CARBOHYD 2509 2513
FT CARBOHYD 2513 2517
FT CARBOHYD 2517 2521
FT CARBOHYD 2521 2525
FT CARBOHYD 2525 2529
FT CARBOHYD 2529 2533
FT CARBOHYD 2533 2537
FT CARBOHYD 2537 2541
FT CARBOHYD 2541 2545
FT CARBOHYD 2545 2549
FT CARBOHYD 2549 2553
FT CARBOHYD 2553 2557
FT CARBOHYD 2557 2561
FT CARBOHYD 2561 2565
FT CARBOHYD 2565 2569
FT CARBOHYD 2569 2573
FT CARBOHYD 2573 2577
FT CARBOHYD 2577 2581
FT CARBOHYD 2581 2585
FT CARBOHYD 2585 2589
FT CARBOHYD 2589 2593
FT CARBOHYD 2593 2597
FT CARBOHYD 2597 2601
FT CARBOHYD 2601 2605
FT CARBOHYD 2605 2609
FT CARBOHYD 2609 2613
FT CARBOHYD 2613 2617
FT CARBOHYD 2617 2621
FT CARBOHYD 2621 2625
FT CARBOHYD 2625 2629
FT CARBOHYD 2629 2633
FT CARBOHYD 2633 2637
FT CARBOHYD 2637 2641
FT CARBOHYD 2641 2645
FT CARBOHYD 2645 2649
FT CARBOHYD 2649 2653
FT CARBOHYD 2653 2657
FT CARBOHYD 2657 2661
FT CARBOHYD 2661 2665
FT CARBOHYD 2665 2669
FT CARBOHYD 2669 2673
FT CARBOHYD 2673 2677
FT CARBOHYD 2677 2681
FT CARBOHYD 2681 2685
FT CARBOHYD 2685 2689
FT CARBOHYD 2689 2693
FT CARBOHYD 2693 2697
FT CARBOHYD 2697 2701
FT CARBOHYD 2701 2705
FT CARBOHYD 2705 2709
FT CARBOHYD 2709 2713
FT CARBOHYD 2713 2717
FT CARBOHYD 2717 2721
FT CARBOHYD 2721 2725
FT CARBOHYD 2725 2729
FT CARBOHYD 2729 2733
FT CARBOHYD 2733 2737
FT CARBOHYD 2737 2741
FT CARBOHYD 2741 2745
FT CARBOHYD 2745 2749
FT CARBOHYD 2749 2753
FT CARBOHYD 2753 2757
FT CARBOHYD 2757 2761
FT CARBOHYD 2761 2765
FT CARBOHYD 2765 2769
FT CARBOHYD 2769 2773
FT CARBOHYD 2773 2777
FT CARBOHYD 2777 2781
FT CARBOHYD 2781 2785
FT CARBOHYD 2785 2789
FT CARBOHYD 2789 2793
FT CARBOHYD 2793 2797
FT CARBOHYD 2797 2801
FT CARBOHYD 2801 2805
FT CARBOHYD 2805 2809
FT CARBOHYD 2809 2813
FT CARBOHYD 2813 2817
FT CARBOHYD 2817 2821
FT CARBOHYD 2821 2825
FT CARBOHYD 2825 2829
FT CARBOHYD 2829 2833
FT CARBOHYD 2833 2837
FT CARBOHYD 2837 2841
FT CARBOHYD 2841 2845
FT CARBOHYD 2845 2849
FT CARBOHYD 2849 2853
FT CARBOHYD 2853 2857
FT CARBOHYD 2857 2861
FT CARBOHYD 2861 2865
FT CARBOHYD 2865 2869
FT CARBOHYD 2869 2873
FT CARBOHYD 2873 2877
FT CARBOHYD 2877 2881
FT CARBOHYD 2881 2885
FT CARBOHYD 2885 2889
FT CARBOHYD 2889 2893
FT CARBOHYD 2893 2897
FT CARBOHYD 2897 2901
FT CARBOHYD 2901 2905
FT CARBOHYD 2905 2909
FT CARBOHYD 2909 2913
FT CARBOHYD 2913 2917
FT CARBOHYD 2917 2921
FT CARBOHYD 2921 2925
FT CARBOHYD 2925 2929
FT CARBOHYD 2929 2933
FT CARBOHYD 2933 2937
FT CARBOHYD 2937 2941
FT CARBOHYD 2941 2945
FT CARBOHYD 2945 2949
FT CARBOHYD 2949 2953
FT CARBOHYD 2953 2957
FT CARBOHYD 2957 2961
FT CARBOHYD 2961 2965
FT CARBOHYD 2965 2969
FT CARBOHYD 2969 2973
FT CARBOHYD 2973 2977
FT CARBOHYD 2977 2981
FT CARBOHYD 2981 2985
FT CARBOHYD 2985 2989
FT CARBOHYD 2989 2993
FT CARBOHYD 2993 2997
FT CARBOHYD 2997 3001
FT CARBOHYD 3001 3005
FT CARBOHYD 3005 3009
FT CARBOHYD 3009 3013
FT CARBOHYD 3013 3017
FT CARBOHYD 3017 3021
FT CARBOHYD 3021 3025
FT CARBOHYD 3025 3029
FT CARBOHYD 3029 3033
FT CARBOHYD 3033 3037
FT CARBOHYD 3037 3041
FT CARBOHYD 3041 3045
FT CARBOHYD 3045 3049
FT CARBOHYD 3049 3053
FT CARBOHYD 3053 3057
FT CARBOHYD 3057 3061
FT CARBOHYD 3061 3065
FT CARBOHYD 3065 3069
FT CARBOHYD 3069 3073
FT CARBOHYD 3073 3077
FT CARBOHYD 3077 3081
FT CARBOHYD 3081 3085
FT CARBOHYD 3085 3089
FT CARBOHYD 3089 3093
FT CARBOHYD 3093 3097
FT CARBOHYD 3097 3101
FT CARBOHYD 3101 3105
FT CARBOHYD 3105 3109
FT CARBOHYD 3109 3113
FT CARBOHYD 3113 3117
FT CARBOHYD 3117 3121
FT CARBOHYD 3121 3125
FT CARBOHYD 3125 3129
FT CARBOHYD 3129 3133
FT CARBOHYD 3133 3137
FT CARBOHYD 3137 3141
FT CARBOHYD 3141 3145
FT CARBOHYD 3145 3149
FT CARBOHYD 3149 3153
FT CARBOHYD 3153 3157
FT CARBOHYD 3157 3161
FT CARBOHYD 3161 3165
FT CARBOHYD 3165 3169
FT CARBOHYD 3169 3173
FT CARBOHYD 3173 3177
FT CARBOHYD 3177 3181
FT CARBOHYD 3181 3185
FT CARBOHYD 3185 3189
FT CARBOHYD 3189 3193
FT CARBOHYD 3193 3197
FT CARBOHYD 3197 3201
FT CARBOHYD 3201 3205
FT CARBOHYD 3205 3209
FT CARBOHYD 3209 3213
FT CARBOHYD 3213 3217
FT CARBOHYD 3217 3221
FT CARBOHYD 3221 3225
FT CARBOHYD 3225 3229
FT CARBOHYD 3229 3233
FT CARBOHYD 3233 3237
FT CARBOHYD 3237 3241
FT CARBOHYD 3241 3245
FT CARBOHYD 3245 3249
FT CARBOHYD 3249 3253
FT CARBOHYD 3253 3257
FT CARBOHYD 3257 3261
FT CARBOHYD 3261 3265
FT CARBOHYD 3265 3269
FT CARBOHYD 3269 3273
FT CARBOHYD 3273 3277
FT CARBOHYD 3277 3281
FT CARBOHYD 3281 3285
FT CARBOHYD 3285 3289
FT CARBOHYD 3289 3293
FT CARBOHYD 3293 3297
FT CARBO

```

RESULT 8
ID YD55_MYCTU STANDARD; PRT: 715 AA.
AC Q11025;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RY1355c.
CN RY1355C OR MT1398 OR MTCY02B10.19C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayama L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z75555; CAAG9988.1; -
DR EMBL: AE007012; AAK45661.1; ALT_INTT.
DR TIGR: MT1398; -
DR Tuberculist; RY1355c; -
DR InterPro; IPR000594; Thif_domain.
DR Pfam; PF00899; Thif; 1.
KW Hypothetical protein; Complete proteome.
SQ
SEQUENCE 715 AA; 78181 MW; 455495248A56041C CRC64;

```

```

Query Match 22.8%; Score 55; DB 1; Length 715;
Best Local Similarity 38.1%; Pred. No. 16;
Matches 8; Conservative 7; Mismatches 4; Indels 2; Gaps 1;

```

```

QY 18 PEAPRIMI--LKRPMPMPRR 36
ID 53 POPDPDLLEAKRWAVYPMRR 73
Db

```

```

RESULT 9
AP22_APIME STANDARD; PRT: 144 AA.
AC P35381; P11525; P11526;
DT 01-OCT-1989 (Rel. 12, Created)

```

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Apidaecin precursor, type 22.
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
OC Aculeata; Apoidea; Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93223697; PubMed=8467807;
RA Casteels-Josson K., Capaci T., Casteels P., Tempst P.;
RT "Apidaecin multiprepeptide precursor structure: a putative mechanism for
RT amplification of the insect antibacterial response."
RL EMBO J. 12:1569-1578(1993).
RN [2]
RP SEQUENCE (APIDAECIN IA/IB).
RC TISSUE-Hemolymph;
RX MEDLINE=90005446; PubMed=2676519;
RA Casteels P., Ampe C., Jacobs F., Vaeck M., Tempst P.;
RT "Apidaecins: antibacterial peptides from honeybees."
RL EMBO J. 8:2387-2391(1989).
CC -!- FUNCTION: APIDAECIN HAVE BACTERICIDAL ACTIVITY; PREDOMINANTLY
CC AGAINST GRAM-NEGATIVE BACTERIA. THEY SEEM TO INTERFERE WITH CELL
CC PROPAagation.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72576; CA51168.1; -
DR PIR: S05383; S05383.
DR PIR: S06675; S06675.
DR PIR: S35331; S35331.
DR InterPro; IPR004828; Apidaecin.
DR Pfam; PF08007; Apidaecin; 4.
KW Insect immunity; Antibiotic; Hemolymph; Signal; Multigene family;
KW Cleavage on pair of basic residues; Repeat.
FT SIGNAL 1 19
FT PROPEP 35 42
FT PEPTIDE 43 60
FT PROPEP 63 70
FT PEPTIDE 71 88
FT PROPEP 91 98
FT PEPTIDE 99 116
FT PROPEP 119 126
FT PEPTIDE 127 144
FT PEPTIDE 144 144
SQ SEQUENCE 144 AA; 16539 MW; 6FA1AD74CB77108D CRC64;

```

```

Query Match 22.4%; Score 54; DB 1; Length 144;
Best Local Similarity 64.7%; Pred. No. 4.4;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 6 PMPWRRRRHEAPEAP 22
ID 26 PTRPRLRLREAPPEAP 42
Db

```

```

RESULT 10
SYCL_MYCTU STANDARD; PRT: 469 AA.
ID SYCL_MYCTU
AC P96862;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cysteineyl-tRNA synthetase 1 (EC 6.1.1.16) (Cysteine--tRNA ligase 1)
DE (CysRS 1).
GN CYSS1 OR CYSS OR RV3580C OR MT3686 OR MTCY06G11.27C.
OS Mycobacterium tuberculosis.

```

```

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gass D., Chillingworth T., Connor R.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Squares S., Squares R.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,
RA Kolony J.F., Nelson W.C., Mayhew L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: ATP + L-cysteine + tRNA(Cys) = AMP +
CC diphosphate + L-cysteinyl-tRNA(Cys).
CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC STRONG, TO METHIONYL-TRNA SYNTHETASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z92774; CAB07154.1; -.
DR EMBL: AE007169; AAK48044.1; -.
DR TIGR: M33686; -.
DR TubercuList; RV3580C; -.
DR InterPro: IPR002308; Cys_trna-synt_1a.
DR InterPro: IPR001412; trna-synt_1.
DR Pfam: PF01406; trna-synt_1e; 1.
DR PRINTS: PR00983; TRNASTYNTHCYS.
DR TIGRfams: TIGR00435; cysS; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminocyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 35 45 "HIGH" REGION.
FT SITE 267 271 "KMSKS" REGION.
FT BINDING 270 270 ATP (BY SIMILARITY).
FT CONFLICT 457 457 D -> E (IN REF. 2).
SQ SEQUENCE 469 AA; 51854 MW; 515F9D19482AD0E CRC64;

Query Match 22.4%; Score 54; DB 1; Length 469;
Best Local Similarity 34.5%; Pred. NO. 14;
Matches 10; Conservative 0; Mismatches 7; Indels 12; Gaps 1;

```

```

OY 6 PWMFWRKHE-----APEAP 22
   ||| |  ||
   ||| |  ||
Db 89 PWMEMATHERAFTAYDALVLPSPAEP 117

```

RESULT 11
SYCL_MYCLE

```

ID SYCL_MYCLE STANDARD; PRT; 473 AA.
AC P57930;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cysteinyl-tRNA synthetase I (EC 6.1.1.16) (Cysteine--trna ligase 1)
DE (CysRS 1).
OS CYSS1 OR CYSS OR ML0323.
OS Mycobacterium lepre.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=2112873; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthey S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., McLean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -|- CATALYTIC ACTIVITY: ATP + L-cysteine + tRNA(Cys) = AMP +
CC diphosphate + L-cysteinyl-tRNA(Cys).
CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC STRONG, TO METHIONYL-TRNA SYNTHETASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL583918; CAC29631.1; -.
DR Lepioma; ML0323; -.
DR InterPro: IPR002308; Cys_trna-synt_1a.
DR InterPro: IPR001412; trna-synt_1.
DR Pfam: PF01406; trna-synt_1e; 1.
DR PRINTS: PR00983; TRNASTYNTHCYS.
DR TIGRfams: TIGR00435; cysS; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminocyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 35 45 "HIGH" REGION.
FT SITE 267 271 "KMSKS" REGION.
FT BINDING 270 270 ATP (BY SIMILARITY).
SQ SEQUENCE 473 AA; 52124 MW; 9FD6CF859C69316D CRC64;

Query Match 22.4%; Score 54; DB 1; Length 473;
Best Local Similarity 34.5%; Pred. NO. 14;
Matches 10; Conservative 0; Mismatches 7; Indels 12; Gaps 1;

```

```

OY 6 PWMFWRKHE-----APEAP 22
   ||| |  ||
   ||| |  ||
Db 89 PWMEMATHERAFTAYDALVLPSPAEP 117

RESULT 12
VGL2_CVH22
ID VGL2_CVH22 STANDARD; PRT; 1173 AA.
AC P15423;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein).
 GN S.
 OS Human coronavirus (strain 229E).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=11137;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90264837; PubMed=2345367;
 RA Raabe T., Schelle-prinz B., Siddell S.G.;
 RT "Nucleotide sequence of the gene encoding the spike glycoprotein of
 human coronavirus HCV 229E.";
 RL J. Gen. Virol. 71:1065-1073(1990).
 CC -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
 TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
 AND IN SYNCYTIIUM FORMATION.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X16816; CAA34723.1; -
 DR PIR: A34766; VCIHHC.
 DR InterPro: IPR002551; Corona_S1.
 DR InterPro: IPR002552; Corona_S2.
 DR Pfam: PF01600; Corona_S1; 1.
 DR Pfam: PF01601; Corona_S2; 1.
 KW Glycoprotein; Envelope protein; Transmembrane; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 1173
 FT DOMAIN 16 1115
 FT TRANSMEM 1116 1135
 FT DOMAIN 1136 1173
 FT DOMAIN 1136 1157
 FT CARBOHYD 23 23
 FT CARBOHYD 62 62
 FT CARBOHYD 98 98
 FT CARBOHYD 147 147
 FT CARBOHYD 171 171
 FT CARBOHYD 176 176
 FT CARBOHYD 220 220
 FT CARBOHYD 243 243
 FT CARBOHYD 336 336
 FT CARBOHYD 333 333
 FT CARBOHYD 440 440
 FT CARBOHYD 464 464
 FT CARBOHYD 518 518
 FT CARBOHYD 538 538
 FT CARBOHYD 542 542
 FT CARBOHYD 568 568
 FT CARBOHYD 581 581
 FT CARBOHYD 587 587
 FT CARBOHYD 663 663
 FT CARBOHYD 671 671
 FT CARBOHYD 930 930
 FT CARBOHYD 1015 1015
 FT CARBOHYD 1020 1020
 FT CARBOHYD 1037 1037
 FT CARBOHYD 1049 1049
 FT CARBOHYD 1061 1061
 FT CARBOHYD 1066 1066
 FT CARBOHYD 1076 1076
 FT CARBOHYD 1082 1082
 FT CARBOHYD 1096 1096
 FT CARBOHYD 1173 AA; 128639 MW; B9CA9A41A796B3BD CMC64;
 SO SEQUENCE

Query Match 22.4%; Score 54; DB 1; Length 1173;
 Best Local Similarity 85.7%; Pred. No. 35;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 4 KRPWWPW 10
 Db 1113 KRPWWPW 1119
 RESULT 13
 ID COR1_HUMAN STANDARD; PRT; 1042 AA.
 AC Q9Y5Q5; Q9UHY2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
 DE converting enzyme) (Corin) (Heart specific serine proteinase ATC2).
 GN CRN.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Heart;
 RX MEDLINE=99262646; PubMed=10329693;
 RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
 RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
 RT from human heart.";
 RL J. Biol. Chem. 274:14926-14935(1999).
 RN [2]
 RP SEQUENCE OF 734-1040 FROM N.A.
 RC TISSUE-Heart;
 RX MEDLINE=20354769; PubMed=11082206;
 RA Hooper J.D., Scaram A.L., Clarke B.E., Normyle J.F., Antal S.T.M.;
 RT "Localization of the mosaic transmembrane serine protease corin to
 RT heart myocytes.";
 RL Eur. J. Biochem. 267:6931-6937(2000).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=20359740; PubMed=10880574;
 RA Yan W., Wu F., Morser J., Wu Q.;
 RT "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
 RT natriuretic peptide-converting enzyme.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529(2000).
 CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
 CC BETWEEN ARG-123 AND SER-124.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART. EXPRESSED IN HEART
 CC MYOCYTES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF133845; AAD31850.1; -
 CC EMBL: AF113248; AAF21966.1; -
 DR HSSP: P00763; IDPO.
 DR MEMOPS: S01.019; -
 DR MIM: 605236; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000024; Fz_domain
 DR InterPro: IPR002172; LDL_receptl_A.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR001190; Strc_receptor.
 DR Pfam: PF00057; ldl_receptl_a; 6.

DR Pfam: PF00089: trypsin; 1.
 DR Pfam: PF01392: Fz; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR SMART: SM00063; FRI; 2.
 DR SMART: SM00192; LDLA; 7.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; Tryp-Spc; 1.
 DR PROSITE: PS01209; LDLRA_1; 6.
 DR PROSITE: PS50068; LDLRA_2; 7.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; FALSE_NEG.
 DR Hydrolase: Serine protease; Transmembrane; Signal-anchor;
 KM Glycoprotein; Repeat.
 FT DOMAIN 1 45
 FT TRANSMEM 46 66
 FT DOMAIN 67 1042
 FT DOMAIN 134 259
 FT DOMAIN 268 304
 FT DOMAIN 305 340
 FT DOMAIN 341 377
 FT DOMAIN 378 415
 FT DOMAIN 450 573
 FT DOMAIN 579 614
 FT DOMAIN 615 653
 FT DOMAIN 654 690
 FT DOMAIN 690 766
 FT DOMAIN 802 1042
 FT ACT_SITE 843 883
 FT ACT_SITE 892 892
 FT ACT_SITE 985 985
 FT ACT_SITE 985 985
 FT DISULFID 790 912
 FT DISULFID 828 844
 FT DISULFID 955 970
 FT DISULFID 981 1010
 FT CARBOHYD 80 80
 FT CARBOHYD 104 104
 FT CARBOHYD 135 135
 FT CARBOHYD 141 141
 FT CARBOHYD 231 231
 FT CARBOHYD 245 245
 FT CARBOHYD 251 251
 FT CARBOHYD 305 305
 FT CARBOHYD 320 320
 FT CARBOHYD 320 320
 FT CARBOHYD 376 376
 FT CARBOHYD 413 413
 FT CARBOHYD 446 446
 FT CARBOHYD 451 451
 FT CARBOHYD 469 469
 FT CARBOHYD 567 567
 FT CARBOHYD 651 651
 FT CARBOHYD 697 697
 FT CARBOHYD 761 761
 FT CARBOHYD 1022 1022
 FT MUTAGEN 985 985
 FT CONFLICT 854 854
 FT CONFLICT 876 876
 FT SEQUENCE 1042 AA: 116564 MW: 7705398EBB60AD2 CRC64;

Query Match 22.2%; Score 53.5; DB 1; Length 1042;
 Best Local Similarity 36.0%; Pred. No. 36;
 Matches 9; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

QY 8 WPMRRKHEAPEAE---PIMILKKW 29
 DB 813 WPMOSLOSSEPSGHICGVLIKKW 837

RESULT 14
 CORI_MOUSE
 ID CORI_MOUSE STANDARD; PRT; 1113 AA.
 AC 092319;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
 DE converting enzyme) (Corin) (low density lipoprotein receptor related
 DE protein 4).
 GN CRN OR LRPA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX MEDLINE=98429596; PubMed=9756624;
 RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;
 RT "A novel low-density lipoprotein receptor-related protein with type II
 RT membrane protein-like structure is abundant in heart."
 RL J. Biochem. 124:784-789(1998).
 CC - FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
 CC BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: Type II membrane protein.
 CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC - SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
 CC - SIMILARITY: CONTAINS 2 FRITZLED (FZ) DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB013874; BAA34371.1; -
 CC HSSP: P00763; IDPO.
 DR MEROPS: S01.019; -
 DR MGD: MGI:1349451; Lrp4.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000024; Fz_domain.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR001190; Srcr_receptor.
 DR Pfam: PF00057; ldl_recept_a; 6.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF01392; Fz; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR SMART: SM00063; FRI; 2.
 DR SMART: SM00192; LDLA; 7.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; Tryp-Spc; 1.
 DR PROSITE: PS50038; Fz; 2.
 DR PROSITE: PS01209; LDLRA_1; 6.
 DR PROSITE: PS50068; LDLRA_2; 7.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR Hydrolase: Serine protease; Transmembrane; Signal-anchor;
 KM Glycoprotein; Repeat.
 FT DOMAIN 1 112
 FT TRANSMEM 113 133
 FT DOMAIN 134 1113
 FT DOMAIN 201 327
 FT DOMAIN 336 372
 FT LDL-RECEPTOR CLASS A 1.

```

FT DOMAIN 373 408 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 409 445 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 446 483 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 518 641 Fz 2.
FT DOMAIN 647 682 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 683 721 LDL-RECEPTOR CLASS A 6.
FT DOMAIN 722 757 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 758 853 SRCR.
FT DOMAIN 869 1113 SERINE PROTEASE.
FT ACT SITE 910 910 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 959 959 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1052 1052 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 857 979 BY SIMILARITY.
FT DISULFID 895 911 BY SIMILARITY.
FT DISULFID 1022 1037 BY SIMILARITY.
FT DISULFID 1048 1077 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 719 719 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1113 AA; 122984 MW; B845B2C5F20DD8EC CRC64;

```

```

Query Match 22.28; Score 53.5; DB 1; Length 1113;
Best Local Similarity 36.0%; Pred. No. 38;
Matches 9; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

```

```

Oy 8 WPMWRKHEAPEAE---PIMILKWK 29
Db 880 WPMWCSLQSEPSGHICCVLIKKW 904

```

```

RESULT 15
FD6C SOYBN STANDARD; PRT; 424 AA.
AC P48628;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=94345008; PubMed=8066133;
RA Hiltz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Stecca K.L.,
RA Yadev N.S.;
RT "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT and its expression in a cyanobacterium.";
RL Plant Physiol. 105:635-641(1994).
CC - FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC - PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC - SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).

```

```

CC - DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/OR BE INVOLVED IN METAL ION BINDING.
CC - SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L29215; AAA50158.1; -
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR Prodom; PD001081; FA_desaturase; 1.
KW Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW Transit peptide.
FT TRANSIT 1 63 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 64 424 OMEGA-6 FATTY ACID DESATURASE.
FT DOMAIN 165 169 HISTIDINE BOX-1.
FT DOMAIN 201 205 HISTIDINE BOX-2.
FT DOMAIN 361 365 HISTIDINE BOX-3.
SQ SEQUENCE 424 AA; 49641 MW; 4F3DF52D4B1A2009 CRC64;

```

```

Query Match 22.08; Score 53; DB 1; Length 424;
Best Local Similarity 34.4%; Pred. No. 17;
Matches 11; Conservative 7; Mismatches 12; Indels 2; Gaps 2;

```

```

Oy 5 WPMWPMWRKHEAPEAEPIMLKWKMPWP-WR 35
Db 192 YPEPMPWRFKHDRH-HAKTNMLREDTAMHPVWK 222

```

```

Search completed: January 15, 2003, 18:08:53
Job time : 13 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:07:41 ; Search time 18 Seconds
(without alignments)
197.610 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKWPMPWRRKheapeapepimilkwpwprrk 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	32.4	144	1 JC1222	indolicidin precursor
2	63	26.1	192	2 H86543	hypothetical prote
3	63	26.1	192	2 D72081	conserved hypotet
4	62	25.7	314	2 S43916	hypothetical prote
5	60	24.9	381	2 B87470	hypothetical prote
6	59.5	24.7	114	2 T36208	hypothetical prote
7	59.5	24.7	2290	1 GNNYE	genome polyprotein
8	59	24.5	95	2 E86447	protein F5D14.5 [1
9	58	24.5	485	2 S74708	hypothetical prote
10	58	24.1	107	2 T35634	hypothetical prote
11	58	24.1	513	2 S21976	probable RNA-direc
12	57.5	23.9	376	2 T40591	hypothetical prote
13	57	23.7	711	2 C40046	hypothetical prote
14	56	23.2	452	2 T28094	antibiotic transpo
15	55.5	23.0	527	2 S33068	hypothetical prote
16	55.5	23.0	990	1 G46335	env polyprotein pr
17	55.5	23.0	1940	2 A59287	myosin heavy chain
18	55	22.8	245	2 C69459	transcription regu
19	55	22.8	352	2 S77448	hypothetical prote
20	55	22.8	715	2 B70741	probable moey' prot
21	54.5	22.6	134	2 E72532	hypothetical prote
22	54.5	22.6	340	2 PH0217	reverse transcript
23	54.5	22.6	451	2 S30401	hypothetical prote
24	54.5	22.6	547	2 T45635	hypothetical prote
25	54	22.4	144	2 S35331	apiadecin 22 precu
26	54	22.4	187	2 AB1647	hypothetical prote
27	54	22.4	469	2 B70607	probable cys5 prot
28	54	22.4	473	2 C86949	probable cysteinyl
29	54	22.4	1173	1 VG1HHC	E2 glycoprotein pr

30	53.5	22.2	538	2 B84759	hypothetical prote
31	53.5	22.2	1113	2 JE0315	low-density lipopr
32	53	22.0	68	2 E82799	hypothetical prote
33	53	22.0	406	2 H69143	coenzyme F420-redu
34	53	22.0	424	2 T07742	omega-6 desaturase
35	53	22.0	480	2 JC7552	Shb-like adapter p
36	53	22.0	691	2 D71430	hypothetical prote
37	53	22.0	949	2 E75352	glycine cleavage s
38	53	22.0	2292	2 S35961	capsid polyprotein
39	52.5	21.8	55	2 T11026	H+-transporting tw
40	52.5	21.8	228	2 AD3326	chloramphenicol O-
41	52.5	21.8	449	2 C84618	hypothetical prote
42	52.5	21.8	471	2 T50016	transcription fact
43	52.5	21.8	490	2 T21365	hypothetical prote
44	52.5	21.8	518	2 T29589	hypothetical prote
45	52.5	21.8	621	2 S37664	peptomeric polypro

ALIGNMENTS

RESULT 1
JC1222
indolicidin precursor - bovine
N:Alternate names: antimicrobial peptide
C:Species: Bos primigenius taurus (cattle)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: JC1222; A42387; S25664
R:del Sal, G.; Storici, P.; Schneider, C.; Romeo, D.; Zanetti, M.
Biochem. Biophys. Res. Commun. 187, 467-472, 1992
A/Title: CDNA cloning of the neutrophil bactericidal peptide indolicidin.
A/Reference number: JC1222; PMID:92392368; PMID:1520337
A/Accession: JC1222
A:Molecule type: mRNA
A:Residues: 1-144 <SAL>
A/Cross-references: EMBL:X67340; NID:9462; PIDN:CAA47755.1; PID:9463
A/Experimental source: bone marrow
R:Selsted, M.E.; Novotny, M.J.; Morris, W.L.; Tang, Y.Q.; Smith, W.; Cullor, J.S.
J. Biol. Chem. 267, 4292-4295, 1992
A/Title: Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils.
A/Reference number: A42387; PMID:92165771; PMID:1537821
A/Accession: A42387
A:Molecule type: protein
A:Residues: 131-143 <SEL>
A/Experimental source: neutrophils
A/Note: Sequence extracted from NCBI backbone (NCBI:83840)
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-143/Product: indolicidin #status experimental <MAT>
F:143/Modified site: amidated carboxyl end (Arg) (amide in mature form from following
Query Match 32.4%; Score 78; DB 1; Length 144;
Best Local Similarity 55.6%; Pred. No. 0.011;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 19 EAEPIMILKWPMPWR 36
DB 126 ELQSVILPKWMPWRR 143

RESULT 2
H86543
hypothetical protein CFJ0426 [imported] - Chlamydomonas pneumoniae (strain J138)
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C/Accession: H86543
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A/Reference number: AB6491; PMID:20330349; PMID:10871362

A;Note: the authors translated the initiation codon GTG for residue 1 as Val
A;Cross-references: EMBL:L13418; NID:g4336964; PIDN:AAA71981.1; PID:g4336967

EC 3.4.12.1; RNA-directed RNA polymerase (EC 2.7.7.48)
 C:Species: encephalomyocarditis virus, EMCV
 A:Note: host Homo sapiens (man)

C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 16-Jul-1999
 C:Accession: A03906; JN0383
 R:Palmerberg, A.C.; Kirby, E.M.; Janda, M.R.; Drake, N.L.; Duke, G.M.; Potratz, K.F.; Co
 Nucleic Acids Res. 12, 2969-2985, 1984
 A:Title: The nucleotide and deduced amino acid sequences of the encephalomyocarditis vir
 A:Reference number: A03906; MUID:84169586; PMID:6324136
 A:Accession: A03906
 A:Molecule type: genomic RNA
 A:Residues: 1-2290 <PAL>
 A:Cross-references: GB:X00463; NID:g61034; PIDN:CAA25152.1; PID:g61035
 R:Petkov, N.A.; Chizhikov, V.E.; Blinov, V.M.; Katginov, V.A.; Mikryukov, N.N.; Gutorov,
 Bioorg. Khim. 10, 274-279, 1984
 A:Title: Nucleotide sequence of the 3'-terminus of encephalomyocarditis virus RNA.
 A:Reference number: JN0383; MUID:85022788; PMID:6091680
 A:Accession: JN0383
 A:Molecule type: genomic RNA
 A:Residues: 1337-1396, 'A', 1519-1517, 'A', 1519-1536, 'E', 1538-1556, 'S', 1558-1611, 'T', 1613-1
 A:Cross-references: GB:M54935
 A:Note: the authors translated the codon CAU for residue 713 as Thr and AAC for residue
 C:Superfamily: foot-and-mouth disease virus genome polyprotein
 C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyltra
 F:1-67/Domain: leader peptide #status predicted <LDP>
 F:66-136/Product: coat protein VP4 #status predicted <VPA>
 F:137-391/Product: coat protein VP2 #status predicted <VP2>
 F:392-622/Product: coat protein VP3 #status predicted <VP3>
 F:623-910/Product: coat protein VP1 #status predicted <VP1>
 F:911-1056/Product: core protein P2-A #status predicted <P2A>
 F:1057-1197/Product: core protein P2-B #status predicted <P2B>
 F:1193-1517/Product: core protein P2-C #status predicted <P2C>
 F:1518-1605/Product: core protein P3-A #status predicted <P3A>
 F:1606-1625/Product: genome-linked protein VPg #status predicted <P3A>
 F:1626-1830/Product: genome-linked protein VPg #status predicted <P3A>
 F:1831-2290/Product: RNA-directed RNA polymerase #status predicted <RDP>

Query Match 24.7% Score 59.5; DB 1; Length 2290;
 Best Local Similarity 26.2%; Pred. No. 38;
 Matches 11; Conservative 6; Mismatches 8; Indels 17; Gaps 1;
 Oy 6 PMPMPRRKHEAPEAPIM-----MLKKP 30
 Db 967 PMPMPRRKHEAPEAPIM-----MLKKP 1008

RESULT 8
 E86447
 Protein F5D14.5 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86447
 R:Phellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;
 anson, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: E86447
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-95 <STO>
 A:Cross-references: GB:AE005172; NID:g8920603; PIDN:AAF81325.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F5D14.5
 A:Map position: 1

Query Match 24.5% Score 59; DB 2; Length 95;
 Best Local Similarity 21.3%; Pred. No. 1.6;
 Matches 10; Conservative 5; Mismatches 4; Indels 28; Gaps 3;

Oy 5 WP-----WMPMPRRKHEAPEAPIMILK-----KMPMPW 34
 Db 46 WPVVVVVAVGAGGGRNMMW-----PVLITTDVGGESMMW 81

RESULT 9
 S74708
 hypothetical protein slr1306 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Varley, PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S74708
 R:Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
 O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S74708
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <KAN>
 A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PIDN:BA16859.1; PID:d101
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 24.5% Score 59; DB 2; Length 485;
 Best Local Similarity 29.3%; Pred. No. 8.6;
 Matches 12; Conservative 6; Mismatches 5; Indels 18; Gaps 2;
 Oy 6 PW-----WMPMPRRKHEAPEAPIM---ILKK 28
 Db 41 PWDGMLALSLGLVTRMRRRHARPEOKOMLTPVLOK 81

RESULT 10
 T35634
 hypothetical protein SC6G9.38 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35634
 R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
 submitted to the EMBL Data Library, June 1999
 A:Reference number: 221584
 A:Accession: T35634
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-107 <SEE>
 A:Cross-references: EMBL:AL079356; PIDN:CA845629.1; GSPDB:GN00070; SCORDB:SC6G9.38
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCORDB:SC6G9.38

Query Match 24.1% Score 58; DB 2; Length 107;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 Oy 3 KKMPMPRRKHEAPEAP 22
 Db 79 KKMPMPRRKHEAPEAP 98

RESULT 11
 S21976
 Probable RNA-directed DNA polymerase (EC 2.7.7.49) (clone MH2C) - rat retrotransposon
 N:Alternate names: reverse transcriptase
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 19-Mar-1997 #sequence_revision 17-Oct-1997 #text_change 26-Aug-1999
 C:Accession: S21976
 R:Kahre, O.; Ilves, H.; Speck, M.
 submitted to the EMBL Data Library, August 1991
 A:Reference number: S16783
 A:Accession: S21976
 A:Molecule type: mRNA

A:Residues: 1-513 <KAB>
 A:Cross-references: EMBL:X61295; NID:g56521; PIDN:CAA43593.1; PID:g56522
 C:Experimental source: clone MH2C
 C:Genetics:
 A:Mobile element: retrotransposon L1
 C:Superfamily: pol polyprotein
 C:Keywords: nucleotidyltransferase; polyprotein; reverse transcriptase

Query Match 24.1%; Score 58; DB 2; Length 513;
 Best Local Similarity 39.4%; Pred. No. 12;
 Matches 13; Conservative 4; Mismatches 12; Indels 4; Gaps 2;
 Oy 1 ILKKPMPWPR---RKHEAPEEPIMLLK-KW 29
 Db 373 IFKMKCFMWRATCRMQIDPCISPTKLSKW 405

RESULT 12

T40591
 Hypothetical protein SPBC646.15c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
 C:Accession: T40591
 R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: Z21938

A:Accession: T40591
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-376 <SEE>
 A:Cross-references: EMBL:AL035216; PIDN:CAA22819.1; GSPDB:GN00067; SPDB:SPBC646.15c
 A:Experimental source: strain 972h-, cosmid c646
 C:Genetics:
 A:Gene: SPDB:SPBC646.15c
 A:Map position: 2
 A:Insertions: 49/1; 126/2; 312/2; 350/1
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC646.15c

Query Match 23.9%; Score 57.5; DB 2; Length 376;
 Best Local Similarity 37.0%; Pred. No. 10;
 Matches 10; Conservative 3; Mismatches 7; Indels 7; Gaps 1;
 Oy 8 WPMRRKHEAPEEPIMLLKMPMPW 34
 Db 236 WHMKRKQKSSS-----LKVPRWGPW 255

RESULT 13

C40046
 Antibiotic transport-associated protein actII-3 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Oct-1999
 C:Accession: C40046
 R:Fernandez-Moreno, M.A.; Caballero, J.L.; Hopwood, D.A.; Malpartida, F.
 Cell 66, 769-780, 1991
 A:Title: The act cluster contains regulatory and antibiotic export genes, direct targets
 A:Reference number: A40046; MUID:91347376; PMID:1878971
 A:Accession: C40046
 A:Molecule type: DNA
 A:Residues: 1-711 <FER>
 A:Cross-references: GB:M4683; NID:g153143; PIDN:AAA26691.1; PID:g153146

Query Match 23.7%; Score 57; DB 2; Length 711;
 Best Local Similarity 47.4%; Pred. No. 23;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Oy 1 ILKKPMPWPRRKHEAPEE 19
 Db 332 IFGRVFWPARPKHGTED 350

RESULT 14

T28094

hypothetical protein ZK899.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T28094
 R:Kershaw, J.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: Z20468
 A:Accession: T28094
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-452 <WIL>
 A:Cross-references: EMBL:Z37140; PIDN:CAA85502.1; GSPDB:GN00028; CESP:ZK899.2
 A:Experimental source: clone ZK899
 C:Genetics:
 A:Gene: CESP:ZK899.2
 A:Map position: X
 A:Insertions: 34/3; 143/2; 227/2; 262/3; 380/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK899.2

Query Match 23.2%; Score 56; DB 2; Length 452;
 Best Local Similarity 32.3%; Pred. No. 19;
 Matches 10; Conservative 3; Mismatches 10; Indels 8; Gaps 2;

Oy 4 KPMWPRRKHEAPEEPIMLLKMPMPW 34
 Db 172 KLVWMTW---HDTDPN----IFDRNMVW 194

RESULT 15

S33068
 myosin heavy chain - fluke (Schistosoma mansoni) (fragment)
 N:Alternate names: surface antigen, 200K
 C:Species: Schistosoma mansoni
 C:Date: 22-Nov-1993 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998
 C:Accession: S33068
 R:Sousson, L.M.A.; Masterson, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.
 J. Immunol. 149, 3612-3620, 1992
 A:Title: Induction of protective immunity in mice using a 62-kDa recombinant fragment
 A:Reference number: A46514; MUID:93056536; PMID:1431131
 A:Accession: S33068
 A:Molecule type: mRNA
 A:Residues: 1-527 <SOI>
 A:Cross-references: EMBL:X65591
 A:Note: the authors translated the codon CAA for residue 346 as Lys
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: ATP; surface antigen

Query Match 23.0%; Score 55.5; DB 2; Length 527;
 Best Local Similarity 30.8%; Pred. No. 25;
 Matches 12; Conservative 7; Mismatches 9; Indels 11; Gaps 2;

Oy 1 ILKKPMPW-----PW---RRKHEAPEEPIMLLK 28
 Db 106 VLNNPMPWRLYTKVKPMLNARQEEEMKKAPELAKKE 144

Search completed: January 15, 2003, 18:09:54
 Job time: 21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:09:36 : Search time 11 Seconds
(without alignments)
66.864 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKWPMPWRKRRKheapeapimllkkwmpwrirk 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141.5	58.7	28	10	US-09-030-619-104
2	137.5	57.1	28	10	US-09-030-619-50
3	99	41.1	21	10	US-09-030-619-69
4	98	40.7	21	10	US-09-030-619-52
5	93.5	38.8	20	10	US-09-030-619-95
6	91	37.8	13	10	US-09-030-619-99
7	91	37.8	13	10	US-09-030-619-72
8	91	37.8	14	10	US-09-030-619-108
9	91	37.8	14	10	US-09-030-619-47
10	87.5	36.3	21	10	US-09-030-619-67
11	87	36.1	12	10	US-09-030-619-112
12	86	35.7	13	10	US-09-030-619-53
13	86	35.7	13	10	US-09-030-619-107
14	86	35.7	13	10	US-09-030-619-109
15	86	35.7	14	10	US-09-030-619-34
16	86	35.7	14	10	US-09-030-619-110
17	86	35.7	12	10	US-09-030-619-73
18	85	35.3	13	10	US-09-030-619-41
19	85	35.3	13	10	US-09-030-619-41

20	85	35.3	13	10	US-09-030-619-103	Sequence 103, App
21	83.5	34.6	15	10	US-09-030-619-39	Sequence 39, Appl
22	83	34.4	12	10	US-09-030-619-44	Sequence 44, Appl
23	83	34.4	13	10	US-09-030-619-105	Sequence 105, App
24	83	34.4	13	10	US-09-030-619-106	Sequence 106, App
25	82.5	34.2	20	10	US-09-030-619-24	Sequence 24, Appl
26	82	34.0	21	10	US-09-030-619-48	Sequence 48, Appl
27	82	34.0	13	10	US-09-030-619-58	Sequence 58, Appl
28	82	34.0	13	10	US-09-030-619-59	Sequence 59, Appl
29	82	34.0	13	10	US-09-030-619-94	Sequence 94, Appl
30	82	34.0	13	10	US-09-030-619-102	Sequence 102, App
31	81	33.6	11	10	US-09-030-619-79	Sequence 79, Appl
32	81	33.6	11	10	US-09-030-619-114	Sequence 114, App
33	78	32.4	11	10	US-09-030-619-78	Sequence 78, Appl
34	78	32.4	11	10	US-09-030-619-113	Sequence 113, App
35	78	32.4	12	10	US-09-030-619-43	Sequence 43, Appl
36	77.5	32.2	12	10	US-09-030-619-40	Sequence 40, Appl
37	77	32.0	12	10	US-09-030-619-82	Sequence 82, Appl
38	76	31.5	12	10	US-09-030-619-23	Sequence 23, Appl
39	75	31.1	12	10	US-09-030-619-30	Sequence 30, Appl
40	75	31.1	12	10	US-09-030-619-111	Sequence 111, App
41	75	31.1	12	10	US-09-030-619-75	Sequence 75, Appl
42	73.5	30.5	12	10	US-09-030-619-76	Sequence 76, Appl
43	73.5	30.5	9	10	US-09-030-619-80	Sequence 80, Appl
44	73	30.3	10	10	US-09-030-619-81	Sequence 81, Appl
45	73	30.3	10	10	US-09-030-619-81	Sequence 81, Appl

ALIGNMENTS

RESULT 1
US-09-030-619-104
; Sequence 104, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krueger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: Niccol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; TITLE OF INVENTION: WITH ANTIBIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030.619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: Fastseq for Windows Version 3.0
; SEQ ID NO 104
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-104
Query Match 58.7%; Score 141.5; DB 10; Length 28;
Best Local Similarity 58.8%; Pred. No. 6.9e-11;
Matches 20; Conservative 2; Mismatches 1; Indels 11; Caps 1;
QY 4 KWPMPWRKRRKHEAPEAPIMLLKKWMPWRK 37
Db 3 RWPMPWRKRRK-----ILKWPMPWRK 25
RESULT 2
US-09-030-619-50
; Sequence 50, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krueger, Timothy J.

```

; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; TITLE OF INVENTION: WITH ANTIBIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indollicidin Analogue
US-09-030-619-50

Query Match          57.1%; Score 137.5; DB 10; Length 28;
Best Local Similarity 61.8%; Pred. No. 1.9e-10;
Matches 21; Conservative 2; Mismatches 0; Indels 11; Gaps 2;

OY      4  ILKKWPMWPMRRKHEAPEPEIMILKKWPMWPMRRK 37
Db      3  RMPWPMWPMRRK-----MIL-RMPWPMWPMRRK 25

RESULT 3
US-09-030-619-69
; Sequence 69, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; TITLE OF INVENTION: WITH ANTIBIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 69
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indollicidin Analogue
US-09-030-619-69

Query Match          41.1%; Score 99; DB 10; Length 21;
Best Local Similarity 64.3%; Pred. No. 4.7e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY      1  ILKKWPMWPMRRKHEAPEPEIMILKK 28
Db      1  ILKKWPMWPMRRK-----MILKK 18

RESULT 4
US-09-030-619-52
; Sequence 52, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
```

```

; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; TITLE OF INVENTION: WITH ANTIBIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indollicidin Analogue
US-09-030-619-52

Query Match          40.7%; Score 98; DB 10; Length 21;
Best Local Similarity 64.3%; Pred. No. 6.1e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY      1  ILKKWPMWPMRRKHEAPEPEIMILKK 28
Db      1  ILKKWPMWPMRR-----IMILKK 18

RESULT 5
US-09-030-619-51
; Sequence 51, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; TITLE OF INVENTION: WITH ANTIBIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indollicidin Analogue
US-09-030-619-51

Query Match          38.8%; Score 93.5; DB 10; Length 20;
Best Local Similarity 60.7%; Pred. No. 2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

OY      1  ILKKWPMWPMRRKHEAPEPEIMILKK 28
Db      1  ILKKWPMWPMRR-----MILKK 17

RESULT 6
US-09-030-619-95
; Sequence 95, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
```

APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: McNicol, Patricia J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
TITLE OF INVENTION: WITH ANTIBIOTICS
FILE REFERENCE: 660081.406
CURRENT APPLICATION NUMBER: US/09/030.619B
CURRENT FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 232
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 95
LENGTH: 13
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-95

Query Match 37.8%; Score 91; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWRRK 13
|||||
Db 1 ILKKPMPWRRK 13

RESULT 7
US-09-030-619-99
Sequence 99, Application US/09030619B
Patent No. US20020035061A1
GENERAL INFORMATION:
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: McNicol, Patricia J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
TITLE OF INVENTION: WITH ANTIBIOTICS
FILE REFERENCE: 660081.406
CURRENT APPLICATION NUMBER: US/09/030.619B
CURRENT FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 232
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 99
LENGTH: 13
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-99

Query Match 37.8%; Score 91; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWRRK 13
|||||
Db 1 ILKKPMPWRRK 13

RESULT 8
US-09-030-619-72
Sequence 72, Application US/09030619B
Patent No. US20020035061A1
GENERAL INFORMATION:
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
APPLICANT: Fraser, Janet R.

APPLICANT: West, Michael H.P.
APPLICANT: McNicol, Patricia J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
TITLE OF INVENTION: WITH ANTIBIOTICS
FILE REFERENCE: 660081.406
CURRENT APPLICATION NUMBER: US/09/030.619B
CURRENT FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 232
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 72
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Indolicidin Analogue
US-09-030-619-72

Query Match 37.8%; Score 91; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWRRK 13
|||||
Db 1 ILKKPMPWRRK 13

RESULT 9
US-09-030-619-108
Sequence 108, Application US/09030619B
Patent No. US20020035061A1
GENERAL INFORMATION:
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: McNicol, Patricia J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
TITLE OF INVENTION: WITH ANTIBIOTICS
FILE REFERENCE: 660081.406
CURRENT APPLICATION NUMBER: US/09/030.619B
CURRENT FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 232
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 108
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-108

Query Match 37.8%; Score 91; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWRRK 13
|||||
Db 1 ILKKPMPWRRK 13

RESULT 10
US-09-030-619-47
Sequence 47, Application US/09030619B
Patent No. US20020035061A1
GENERAL INFORMATION:
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.

```

; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indolicidin Analogue
US-09-030-619-47

```

```

Query Match          36.3%; Score 87.5; DB 10; Length 21;
Best Local Similarity 60.0%; Pred. No. 0.0001; 0;
Matches 15; Conservative 1; Mismatches 0; Indels 9; Gaps 1;

```

```

QY      4 LKKWPMWPMWRRK 13
        |||||||
Db      3 RMPWPMWPMWRRK-----IMILKK 18

```

```

RESULT 11
US-09-030-619-67
; Sequence 67, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indolicidin Analogue
US-09-030-619-67

```

```

Query Match          36.1%; Score 87; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LKKWPMWPMWRRK 13
        |||||||
Db      1 LKKWPMWPMWRRK 12

```

```

RESULT 12
US-09-030-619-112
; Sequence 112, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.

```

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-112

```

```

Query Match          35.7%; Score 86; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 ILKKPMWPMWRR 12
        |||||||
Db      1 ILKKPMWPMWRR 12

```

```

RESULT 13
US-09-030-619-53
; Sequence 53, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Indolicidin Analogue
US-09-030-619-53

```

```

Query Match          35.7%; Score 86; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 9.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 ILKKPMWPMWRR 12
        |||||||
Db      1 ILKKPMWPMWRR 12

```

```

RESULT 14
US-09-030-619-107
; Sequence 107, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfile, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING

```


Fri Jan 17 19:11:32 2003

```

; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE OF INVENTION: WITH ANTI-BIOTICS
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 107
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-107

```

```

Query Match          35.7%; Score 86; DB 10; Length 13;
Best Local Similarity 41.9%; Pred.No. 9.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 18; Gaps 1;

```

```

OY      7 MWPMRRHAEPEAEPIMLKKMPMPMPRRK 37
      11 |||||||||
DB      1 MW-----KKMPMPMPRRK 13

```

```

RESULT 15
US-09-030-619-109
; Sequence 109, Application US/09030619B
; Patent No. US20020035061A1
; GENERAL INFORMATION:
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Erfle, Douglas
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: McNicol, Patricia J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING CATIONIC PEPTIDES ALONE OR IN COMBINATION
; FILE REFERENCE: 660081.406
; CURRENT APPLICATION NUMBER: US/09/030,619B
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Cationic Peptide Analogue
US-09-030-619-109

```

```

Query Match          35.7%; Score 86; DB 10; Length 13;
Best Local Similarity 100.0%; Pred.No. 9.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 ILKKMPMPMPRR 12
      11 |||||||||
DB      1 ILKKMPMPMPRR 12

```

Search completed: January 15, 2003, 18:13:20
Job time : 11 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:07:56 : Search time 14 Seconds
(without alignments)
77.761 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKKWPMWPMRRKHEAPEPIMLKKWPMWRK 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/laa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/laa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/laa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/laa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/laa/PCCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/laa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133.5	55.4	63	4	US-09-099-631A-12
2	99.5	41.3	16	4	US-08-702-054B-38
3	99	41.1	21	4	US-08-915-314-54
4	98	40.7	21	4	US-08-915-314-56
5	93.5	38.8	20	4	US-08-915-314-55
6	91	37.8	13	4	US-08-915-314-30
7	91	37.8	13	4	US-08-915-314-62
8	91	37.8	13	4	US-08-915-314-64
9	91	37.8	13	4	US-08-915-314-64
10	91	37.8	13	4	US-09-042-071-36
11	91	37.8	14	4	US-08-915-314-57
12	89	36.9	15	4	US-08-702-054B-40
13	87.5	36.3	21	4	US-08-915-314-46
14	87	36.1	12	4	US-08-915-314-52
15	86	35.7	12	4	US-08-915-314-74
16	86	35.7	12	4	US-08-702-054B-5
17	86	35.7	13	4	US-08-915-314-51
18	86	35.7	13	4	US-08-915-314-58
19	86	35.7	13	4	US-08-915-314-58
20	86	35.7	14	4	US-08-702-054B-34
21	85.5	35.5	16	4	US-08-915-314-59
22	85	35.3	12	4	US-08-702-054B-11
23	85	35.3	13	4	US-08-915-314-69
24	85	35.3	13	4	US-08-915-314-38
25	85	35.3	13	4	US-08-915-314-45
26	85	35.3	13	4	US-08-702-054B-1
27	85	35.3	13	4	US-08-702-054B-17

28	83.5	34.6	15	4	US-08-702-054B-39	Sequence 39, Appl
29	83	34.4	12	4	US-08-915-314-24	Sequence 24, Appl
30	83	34.4	13	4	US-08-915-314-49	Sequence 49, Appl
31	83	34.4	13	4	US-08-915-314-50	Sequence 50, Appl
32	83	34.4	13	4	US-08-702-054B-30	Sequence 30, Appl
33	83	34.4	13	4	US-08-702-054B-31	Sequence 31, Appl
34	83	34.4	13	4	US-08-702-054B-35	Sequence 35, Appl
35	83	34.4	20	4	US-08-915-314-47	Sequence 47, Appl
36	82.5	34.2	21	4	US-08-915-314-48	Sequence 48, Appl
37	82	34.0	13	4	US-08-915-314-25	Sequence 25, Appl
38	82	34.0	13	4	US-08-915-314-66	Sequence 66, Appl
39	82	34.0	13	4	US-08-915-314-67	Sequence 67, Appl
40	82	34.0	13	4	US-08-702-054B-33	Sequence 33, Appl
41	81	33.6	11	4	US-08-915-314-75	Sequence 75, Appl
42	80	33.2	14	4	US-08-702-054B-18	Sequence 18, Appl
43	80	33.2	15	4	US-08-702-054B-41	Sequence 41, Appl
44	80	33.2	16	4	US-08-702-054B-2	Sequence 2, Appl
45	79	32.8	17	4	US-08-702-054B-42	Sequence 42, Appl

ALIGNMENTS

```
RESULT 1
US-09-099-631A-12
; Sequence 12, Application US/09099631A
; Patent No. 6444645
; GENERAL INFORMATION:
; APPLICANT: Selssted, Michael E.
; TITLE OF INVENTION: Crosslink-Stabilized Indolicidin Analogs
; FILE REFERENCE: P-UC 3050
; CURRENT APPLICATION NUMBER: US/09/099, 631A
; CURRENT FILING DATE: 1998-06-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-099-631A-12

Query Match      55.4%; Score 133.5; DB 4; Length 63;
Best Local Similarity 64.7%; Pred. No. 2.9e-10;
Matches 22; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY      4  KWPMPWPMRRKHEAPEPIMLKK-KWPMPWPMRR 36
Db      11 KWPMPWPMRRM-----ARIAMILPKWPMWPMRR 38

RESULT 2
US-08-702-054B-38
; Sequence 38, Application US/08702054B
; Patent No. 6191254
; GENERAL INFORMATION:
; APPLICANT: Falls, Timothy J.
; APPLICANT: Hancock, Robert E. W.
; TITLE OF INVENTION: ANTINICHOBIAL CATIONIC PEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,054B
FILING DATE: 23-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,687
FILING DATE: 23-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07420/013001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-702-054B-38

Query Match 41.3%; Score 99.5; DB 4; Length 16;
Best Local Similarity 43.2%; Pred. No. 1.2e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 21; Gaps 1;

QY 1 ILKKPMPWRRKHEAPEAPIMILKKPMPWRRK 37
DB 1 ILKKPMPWRRKHEAPEAPIMILKKPMPWRRK 16

RESULT 3
US-08-915-314-54
Sequence 54, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
US-08-915-314-54

Query Match 41.1%; Score 99; DB 4; Length 21;
Best Local Similarity 64.3%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 ILKKPMPWRRKHEAPEAPIMILKK 28
DB 1 ILKKPMPWRRKHEAPEAPIMILKK 18

RESULT 4
US-08-915-314-56
Sequence 56, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-915-314-56

Query Match 40.7%; Score 98; DB 4; Length 21;
Best Local Similarity 64.3%; Pred. No. 2.5e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 ILKKPMPWRRKHEAPEAPIMILKK 28
DB 1 ILKKPMPWRRKHEAPEAPIMILKK 18

RESULT 5
US-08-915-314-55
Sequence 55, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.

APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-915-314-55

Query Match 38.8%; Score 93.5; DB 4; Length 20;
Best Local Similarity 60.7%; Pred.No. 8.8e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

QY 1 ILKKPMPWRRKHEAPEAEPIMILKK 28
Db 1 ILKKPMPWRR-----MILKK 17

RESULT 6
US-08-915-314-30
Sequence 30, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314

FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-915-314-30

Query Match 37.8%; Score 91; DB 4; Length 13;
Best Local Similarity 100.0%; Pred.No. 1.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWRRK 13
Db 1 ILKKPMPWRRK 13

RESULT 7
US-08-915-314-62
Sequence 62, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
FEATURE: Modified-site
NAME/KEY: 1
LOCATION: 1
OTHER INFORMATION: /note="D-Form of Isoleucine"
US-08-915-314-62

Query Match 37.8%; Score 91; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILKKWPMWPMWRRK 13
DB 1 ILKKWPMWPMWRRK 13

RESULT 8

US-08-915-314-63
; Sequence 63, Application US/08915314
; Patent No. 6180604
; GENERAL INFORMATION:
; APPLICANT: Fraser, Janet R.
; APPLICANT: West, Michael H.P.
; APPLICANT: Krieger, Timothy J.
; APPLICANT: Taylor, Robert
; APPLICANT: Exile, Douglas
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
; TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
; NUMBER OF SEQUENCES: 90
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,314
; FILING DATE: 20-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6180604tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 660081.405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /note= "D-Form of Lysine"
US-08-915-314-63

Query Match 37.8%; Score 91; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILKKWPMWPMWRRK 13
DB 1 ILKKWPMWPMWRRK 13

RESULT 9
US-08-915-314-64
; Sequence 64, Application US/08915314
; Patent No. 6180604
; GENERAL INFORMATION:
; APPLICANT: Fraser, Janet R.

APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Exile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "D-Form of Isoleucine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /note= "D-Form of Lysine"
US-08-915-314-64

Query Match 37.8%; Score 91; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ILKKWPMWPMWRRK 13
DB 1 ILKKWPMWPMWRRK 13

RESULT 10

US-09-042-071-36
; Sequence 36, Application US/09042071
; Patent No. 6294372
; GENERAL INFORMATION:
; APPLICANT: Burian, Jan
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: REPLICATION GENES AND GENE PRODUCTS FROM
; TITLE OF INVENTION: SMALL CRYPTIC PLASMIDS AND METHODS FOR CONSTRUCTING
; TITLE OF INVENTION: CONTROLLED-REPLICATION PLASMID VECTORS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,071
FILING DATE: 13-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 660081.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-042-071-36

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPMPRRK 13
Db 1 ILKKPMPMPRRK 13

RESULT 11
US-08-915-314-57
Sequence 57, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tendburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
US-08-915-314-57

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPMPRRK 13
Db 1 ILKKPMPMPRRK 13

RESULT 12
US-08-702-054B-40
Sequence 40, Application US/08702054B
Patent No. 6191254
GENERAL INFORMATION:
APPLICANT: Falls, Timothy J.
APPLICANT: Hancock, Robert E. W.
APPLICANT: Gough, Monisha
TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
TITLE OF INVENTION: AND METHODS OF SCREENING FOR THE SAME
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
City: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,054B
FILING DATE: 23-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,687
FILING DATE: 23-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07420/013001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-702-054B-40

Query Match
Best Local Similarity 36.9%; Score 89; DB 4; Length 15;
Matches 14; Conservative 0; Mismatches 0; Indels 18; Gaps 1;

QY 1 ILKKPMPMPRRKHEAPEEPIIMILKKPMPW 32
Db 1 ILKKPMPMPRRK-----MW 14

RESULT 13
US-08-915-314-46
Sequence 46, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.

```

APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-915-314-46

Query Match          36.3%; Score 87.5; DB 4; Length 21;
Best Local Similarity 60.0%; Pred. No. 5.2e-05;
Matches 15; Conservative 1; Mismatches 0; Indels 9; Gaps 1;

QY 4 KWPWPMRKHAEPEAEPIMLKK 28
DB 3 KWPWPMRKHAEPEAEPIMLKK 18

RESULT 14
US-08-915-314-52
Sequence 52, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-915-314-74

Query Match          35.7%; Score 86; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-915-314-52

Query Match          36.1%; Score 87; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKKPWPWPMRKK 13
DB 1 LKKPWPWPMRKK 12

RESULT 15
US-08-915-314-74
Sequence 74, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-915-314-74

Query Match          35.7%; Score 86; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


Oy	1	ILKKPMPWRR	12
Db	1	ILKKPMPWRR	12

Search completed: January 15, 2003, 18:10:13
Job time : 14 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2003, 18:04:30 : Search time 35 Seconds
(without alignments)
140.865 Million cell updates/sec

Title: US-09-444-281-35-COPY
Perfect score: 241
Sequence: 1 ILKMPWMPWRRKheapeapeplmlkkwmpwrirk 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141.5	58.7	28	21	AAV91800
2	137.5	57.1	27	19	AAW66363
3	133.5	55.4	63	21	AAW44668
4	133.5	55.4	63	21	AAW57142
5	99.5	41.3	16	18	AAW12899
6	99.5	41.1	21	19	AAW24582
7	99	41.1	21	21	AAV91806
8	98	40.7	21	21	AAW24571
9	98	40.7	21	21	AAV91808
10	93.5	38.8	20	19	AAW24570

11	93.5	38.8	20	21	AAV91807
12	91	37.8	13	18	AAW12873
13	91	37.8	13	19	AAW24609
14	91	37.8	13	19	AAW66378
15	91	37.8	13	19	AAW71690
16	91	37.8	13	21	AAW94495
17	91	37.8	13	21	AAV92795
18	91	37.8	13	21	AAV91773
19	91	37.8	13	21	AAV91774
20	91	37.8	13	21	AAV91818
21	91	37.8	13	21	AAV91819
22	91	37.8	13	21	AAV91820
23	91	37.8	13	23	ABB81254
24	91	37.8	14	19	AAW24583
25	91	37.8	14	19	AAV91811
26	89	36.9	15	18	AAW13802
27	87.5	36.3	21	19	AAW24552
28	87.5	36.3	21	19	AAW66376
29	87.5	36.3	21	21	AAV91796
30	87	36.1	12	19	AAW24580
31	87	36.1	12	21	AAV91804
32	86	35.7	12	18	AAW12877
33	86	35.7	12	19	AAW24615
34	86	35.7	12	21	AAV91833
35	86	35.7	13	18	AAW12896
36	86	35.7	13	19	AAW24613
37	86	35.7	13	19	AAW24572
38	86	35.7	13	21	AAV91803
39	86	35.7	13	21	AAV91812
40	86	35.7	14	21	AAW24573
41	86	35.7	14	21	AAV91813
42	85.5	35.5	16	18	AAW12882
43	85	35.3	12	19	AAW24586
44	85	35.3	12	21	AAV91828
45	85	35.3	13	18	AAW27179

ALIGNMENTS

RESULT 1	AAV91800	AAV91800 standard; Peptide: 28 AA.
ID	AAV91800	
XX	AAV91800:	
AC	06-JUN-2000 (first entry)	
XX		
DE	Amino acid sequence of cationic peptide MBI 11B20CN.	
KW	Cationic peptide; tumour; pharmaceutical composition; cancer; treatment; leukemia; polyoxalkylene-modified; APO; lymphoma; multiple myeloma; breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon; multidrug resistance.	
KW		
XX	Synthetic.	
OS		
XX	WO9965506-A2.	
PN	23-DEC-1999.	
PD		
XX	14-JUN-1999; 99WO-CA00552.	
PF		
XX	12-JUN-1998; 98US-0096541.	
PR	(MICR-) MICROLOGIX BIOTECH INC.	
PA	Friedland HD, Krieger TJ, Taylor R, Erffle D, Fraser JR, West MHP;	
XX	WPI; 2000-223549/19.	
DR		
XX	Novel pharmaceutical composition containing optionally activated polyoxalkylene-modified cationic peptides, useful for treating tumours	
PT		

-
Claim 1; Page 15; 94pp; English.

This sequence represents a cationic peptide amino acid sequence, which can be used in the pharmaceutical composition of the invention. The invention relates to a pharmaceutical composition containing at least one activated polyoxalylene (APO)-modified cationic peptide. The modification of peptides with APO increases their activity against tumour cells, including those with a multidrug resistant phenotype. The pharmaceutical composition can be used to treat tumours, specifically lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary, cervix, uterus, skin, prostate, liver and colon.

Sequence 28 AA:

Query Match 58.7%; Score 141.5; DB 21; Length 28;
Best Local Similarity 58.8%; Pred No. 3.6e-11;
Matches 20; Conservative 2; Mismatches 1; Indels 11; Gaps 1

4 KWPWPPRRKHEDPEAPRIMLLKKWPPWPRRK 37
:||||||| 11:|||||||
3 RWPWPPWPRRK-----ILMRPWWPWRK 25

RESULT 2
AAW66363
ID AAW66363 standard; peptide; 27 AA.
XX AAW66363;
AC
XX
DT 12-JAN-1999 (first entry)
XX
DE Indolicidin analogue MBI 11B20.
XX
KW Indolicidin analogue; resistance; cationic peptide; antibiotic;
KM bacterial infection; tolerance; antibacterial; microorganism;
KN bacteria; fungus; parasite; virus.
XX
OS Bos taurus.
OS Synthetic.
XX
XX
PN WO9840401-A2.
XX
PD 17-SEP-1998.
XX
PF 10-MAR-1998; 98WO-CA00190.
XX
PR 25-FEB-1998; 98US-0030619.
PR 10-MAR-1997; 97US-0040649.
PR 20-AUG-1997; 97US-0915314.
PR 26-SEP-1997; 97US-0060099.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
XX Fraser JR, McNicol PJ, West MMP;
PI WPI; 1998-520800/44.
DR
XX
PT New indolicidin peptide analogues - useful for, e.g. enhancing
XX activity of antibiotic or overcoming tolerance, acquired resistance
PT or inherent resistance of microorganisms
XX
XX
PS Claim 1; Page 91; 105pp; English.
XX
XX The present sequence represents an indolicidin analogue. The present
CC invention describes compositions and methods for treating infection,
CC especially bacterial infections. The compositions and methods use
CC cationic peptides in combination with an antibiotic agent which are
CC then administered to a patient to enhance the activity of the antibiotic
CC agent, to overcome: (a) tolerance; (b) acquired resistance; and (c)
CC inherent resistance. The combinations of antibiotics and cationic
CC peptides can provide synergistic activity against a microorganism that

	CC	is tolerant,	inherently resistant, or has acquired resistance to an antibiotic agent.	They can be used for killing e.g. bacteria, fungi, parasites and viruses.	
	XX	Sequence	27 AA:		
	SQ	Query Match	57.1%; Score 137.5; DB 19;	Length 27;	
		Best Local Similarity	61.8%;	Pred. No. 1.le-10;	
	MATCHES	Matches 21;	Conservative 2;	Mismatches 0;	Indels 11; Gaps 2;
OY		4 KWPMPWPRRKHEAPEPIMLKWKPPWPMWRK	37 :: :		
DG		3 RWPMPWPMWRK-----ML-RWPMPWPMWRK	25		
	RESULT 3				
ID	AAY44668	standard; Protein; 63 AA.			
XX					
AC	AAY44668;				
XX					
DT	18-APR-2000	(first entry)			
XX					
DE	Poly-(Indol (1-13)-Met-Ala-Arg-Ile-Ala-Met)3 protein.				
KX	Crosslinked indolicidin analog; X-indolicidin; poly-Indol 1-13;				
KM	stability; bovine neutrophil; antimicrobial; antibacterial; fungicide;				
KV	protozoacytic; vitucide; anti-HIV; human immunodeficiency virus-1;				
KW	HIV-1; gram positive bacterium; gram negative; Staphylococcus aureus;				
KW	Escherichia coli; Salmonella typhimurium; yeast; fungi; protozoa;				
KM	Candida albicans; Cryptococcus neoformans; Giardia; Acanthamoeba;				
KW	hexapeptide spacer.				
OS	Synthetic.				
XO	Bos sp.				
XX					
FH	Key	Location/Qualifiers			
FT	Region	1..5			
FT		/label= Enterokinase_recognition_site			
FT	Cleavage-site	5..6			
FT		/label= Enterokinase_cleavage_site			
FT	Cleavage-site	6..7			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Cleavage-site	20..21			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Cleavage-site	25..26			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Cleavage-site	39..40			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Cleavage-site	44..45			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Cleavage-site	58..59			
FT		/label= Cyanogen_bromide_cleavage_site			
FT	Region	20..25			
FT		/label= Hexapeptide_spacer			
FT	Region	39..44			
FT		/label= Hexapeptide_spacer			
FT	Region	58..63			
FT		/label= Hexapeptide-spacer			
PN	WO9965510-A1.				
XX					
PD	23-DEC-1999.				
XX					
PF	20-MAY-1999;	99WO-US11165.			
XX					
PR	18-JUN-1998;	98US-0099631.			
XX					
PA	(REGC) UNIV CALIFORNIA.				
XX					
PI	Selectd ME, Osapay K;				
XX					
DP	WPI; 2000-147133/13.				

DR	N-PSDB;AAZ49764.
XX	
PT	Crosslinked indolicidin analogs with antimicrobial activity against
PT	bacteria, yeast, fungi, protozoa and viruses -
XX	
PS	Example 1C; Fig 1; 53pp; English.
XX	
CC	The patent discloses crosslinked analogs of indolicidin (Indol 1-13)
CC	which is a naturally occurring peptide isolated from bovine neutrophils
CC	and has antimicrobial activity. The crosslinked indolicidin
CC	(X-indolicidin) analogs are stable and have antimicrobial activity
CC	against gram positive and negative bacteria (e.g. Staphylococcus aureus,
CC	Escherichia coli and Salmonella typhimurium), yeasts and fungi (e.g.
CC	Candida albicans, Cryptococcus neoformans), protozoa (e.g. Giardia
CC	species and Acanthamoeba species), and viruses (e.g. HIV-1).
CC	They can be used for reducing or inhibiting the growth or survival of
CC	microorganisms in an environment e.g. a food or food product, a
CC	solution, an inanimate object comprising a surface, or a mammal.
CC	The present sequence is a protein comprising three
CC	copies of Indol 1-13 each separated by a hexapeptide spacer sequence.
CC	A recombinant construct encoding this sequence was used for the
CC	expression of Indol-homoserine (Ise) analog. The ability of
CC	Indol-Ise analog to maintain antimicrobial activity provides a means to
CC	produce X-indolicidin analog precursors in sufficient quantities.
XX	
SQ	Sequence 63 AA:
Query Match	55.4%; Score 133.5; DB 21; Length 63;
Best Local Similarity	64.7%; Pred. No. 9e-10; Mismatches
Matches 22; Conservative 0; Indels 7; Gaps 2;	
OY	4 KWPMPWPMRRKHDEPPEAPIMLK-KWPMPWMRR 36
Db	11 KWPMPWMRRM-----ARIAMILPKMKWPMPWMRR 38
RESULT 4	
AAV57142	
AAV57142	standard; Protein: 63 AA.
XX	
AC	AAV57142;
XX	
DT	28-FEB-2000 (first entry)
XX	
DE	Indolicidin fusion peptide amino acid sequence.
XX	
KW	Indolicidin analogue; antimicrobial activity; helminth; bacteria; virus;
KW	treatment; inhibit growth; micro-organism; contact lens solution;
KW	transgenic plant; surgical instrument; yeast; fungi; protozoa.
OS	Synthetic.
XX	
PN	WO9958141-A1.
XX	
PD	18-NOV-1999.
XX	
PF	05-MAY-1999; 99WO-US09942.
XX	
PR	12-MAY-1998; 98US-0076227.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	
P1	Seleted ME;
XX	
DR	WPI: 2000-053028/04.
XX	
DR	N-PSDB; AAZ45123.
XX	
PT	New indolicidin analogues, active against bacteria, yeast, fungi,
PT	protozoa and virus, used for, e.g. treating infections -
XX	
PS	Disclosure; Fig 6; 62pp; English.
XX	
CC	This is the amino acid sequence of an example of a fusion protein which

CC	consists of an indolicidin analogue linked to another peptide.
CC	Peptides AAY5109-Y51138 and AAY51143-Y51144 are new indolicidin
CC	analogues, which have a homoserine residue and/or a truncated amino
CC	terminal region. The analogues have the following amino acid sequence:
CC	Xaa1-Xaa2-Xaa3-Xaa4-Xaa5-Xaa6-Pro-Xaa6-Xaa6-Pro-Xaa6-Xaa7-Xaa7-Xaa8
CC	Where:
CC	Xaa1 = Ile, Leu, Val, Ala, Gly or absent;
CC	Xaa2 = Ile, Leu, Val, Ala, Gly or absent;
CC	Xaa3 = Pro or absent;
CC	Xaa4 = Trp, Phe or absent;
CC	Xaa5 = Arg, Lys or absent;
CC	Xaa6 = Trp or Phe;
CC	Xaa7 = Arg, Lys or absent;
CC	Xaa8 = homoserine (Hse), Met, Met-Xaa9-Met or absent, and
CC	Xaa9 = at least one amino acid;
CC	provided that if Xaa1 is present, Xaa8 = Hse, Met or Met-Xaa9-Met;
CC	and further provided that: if Xaa2 is absent, Xaa1 is absent; if Xaa3 is
CC	absent, Xaa1 and Xaa2 are absent; if Xaa4 is absent, Xaa1, Xaa2 and Xaa3
CC	are absent; and if Xaa5 is absent, Xaa1, Xaa2, Xaa3 and Xaa4 are absent.
CC	The indolicidin analogues can be used to create a fusion polypeptide
CC	consisting of the analogue linked to a peptide. The indolicidin
CC	analogues have antimicrobial activity against gram positive bacteria,
CC	gram negative bacteria, yeast, fungus, protozoa and viruses (e.g. HIV-1).
CC	They are also active against helminths. The analogues can be used for
CC	reducing or inhibiting growth or survival of a microorganism. They can be
CC	used for treating infections. They can also be included in a liquid such
CC	as water or an aqueous solution, e.g. contact lens solution. The
CC	analogues have potential uses in food products, and in objects such as
CC	the surface of an instrument used to prepare food or to perform surgery.
CC	Transgenic plants or animals useful in the food industry can be produced
CC	by introducing a nucleic acid molecule encoding an indolicidin analogue
CC	into the germ-line cells of such organisms.
XX	
SO	Sequence 63 AA;
QY	4 KMPWMPWRRKHEAPEEPIITLK-KMPWMPWRR 36
Db	11 KMPWMPWRRM-----ARIATILPKMPWMPWRR 38
RESULT 5	
AAWI2899	
ID	AAWI2899 standard; peptide; 16 AA.
XX	
AC	AAWI2899;
XX	
DT	10-DEC-1997 (first entry)
XX	
DE	Antimicrobial cationic peptide CP-26.
XX	
KW	Bacterial; viral; antitumour; food; preservative; inhibitor; growth;
KW	bacterium; yeast; endotoxaemia; sepsis; antibiotic; fungal;
KW	antiviral; Candida albicans; steriliant; Salmonella; Yersinia;
KW	Shigella.
XX	
OS	Synthetic.
XX	
PN	W09708199-A2.
PD	06-MAR-1997.
XX	
PF	23-AUG-1996; 96WO-IB00996.
XX	
PR	23-AUG-1995; 95US-0002687.
XX	
PA	(UYBR-) UNIV BRITISH COLUMBIA.
XX	
PI	Falla TJ, Gough M, Hancock RW;
XX	

DR MPI: 1997-179179/16.

XX Cationic peptide(s) having anti-microbial activity - used for the

PT inhibition of bacterial and viral growth, as an antitumour agent,

PT and as a food preservative

XX

PS Claim 3; Page 66; 89pp; English.

XX The present sequence represents a specifically claimed novel isolated

CC cationic peptide which has antimicrobial activity. The amino acid

CC sequence of antimicrobial cationic peptides (including the present

CC sequence) is selected from: X1X1Prox2X3X2Pro(X2X2Pro)(X2X3(X5)O;

CC X1X1Prox2X3X4(X5)Prox2X3X3; X1X1X3(ProTrp)X3X2X5X2X5X2(X5)O;

CC X1X1X3X3X2Pro(X2X2Pro)(X2(X5)m; where m = 1-5; n = 1-2; o = 2-5; r

CC = 0-8; u = 0-1; X1 = Ile, Leu, Val, Phe, Tyr, Trp or Met; X2 = Trp or

CC Phe; X3 = Arg or Lys; X4 = Trp or Lys; and X5 = Phe, Trp, Arg, Lys or

CC Pro. The peptides are preferably amidated or carbonylmethylated. The

CC peptides may be used in methods for inhibiting the growth of a bacterium

CC or yeast, or for inhibiting an endotoxaemia or sepsis associated

CC disorder in a subject. The peptides have a broad activity against

CC antibiotic resistant bacteria, combined with activity against the

CC medically important fungus *Candida albicans*. In addition, the peptides

CC are useful as antitumour agents and/or antiviral agents. The peptides

CC may be used as sterilants or preservatives of materials susceptible to

CC microbial or viral contamination, e.g. in processed foods to inhibit

CC *Salmonella*, *Yersinia* and *Shigella*. The peptides are compact and tend to

CC have a unique polypoline type II extended helix structure that permits

CC them to span the membrane with relatively few amino acids. The peptides

CC possess the ability to work synergistically with antibiotics, and in

CC addition, some of them possess anti-endotoxin activity.

XX

SQ Sequence 16 AA;

Query Match 41.3%; Score 99.5; DB 18; Length 16;

Best Local Similarity 43.2%; Pred. No. 3.7e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 21; Gaps 1;

OY 1 ILKKPMPWMPRRKHEAPEAPIMILKKPMPWMPRRK 37

Db 1 ILKKMPW-----WPMWMPRRK 16

RESULT 6

AA24582

ID AAY24582 standard; peptide; 21 AA.

XX AAY24582;

AC 18-AUG-1999 (first entry)

XX

DT Indolicidin analogue #34.

XX

DE Indolicidin: bacterial infection; photo-oxidised solubilisier;

XX anti-microbial; antibiotic; antilarylmuc; surface disinfectant;

KW additive; shampoo; soap; insecticide; herbicide; preservative;

KW food; technical material.

XX

OS Synthetic.

XX

XX WO9807745-A2.

PN

XX

PD 26-FEB-1998.

XX

PF 21-AUG-1997; 97WO-US14779.

XX

PR 13-JAN-1997; 97US-0034949.

XX

PR 21-AUG-1996; 96US-0024754.

XX

PA (MICR-) MICROLOGIX BIOTECH INC.

XX

PI Erfle D, Fraser JR, Krieger TJ, Taylor R, West MH;

XX

DR MPI; 1998-169090/15.

XX New indolicidin analogues with antimicrobial activity and related

PT nucleic acid - vectors, transformed cells and antibodies, also

PT conjugates with polyoxalkylene glycol and fatty acid to reduce

PT toxicity, useful therapeutically, as disinfectants etc.

XX

PS Claim 13; Page 89; 129pp; English.

XX AAY24549 to AAY24615 represent indolicidin analogues of formulae

CC (I)-(VIII) containing up to 25 amino acids (aa): RXXXXZXB (I), BXZXXZXB

CC (II), BBBXXZXB (III), BXXZXXBBB(AA)nMILBAGS (IV), BXXZXB(AA)nM

CC (V), LBBnXnXZnXnR (VI), LKnZXXZXRK (VII) and BXXZXXZBBB (VIII).

CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,

CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;

CC in (VIII) at least 2 X = F or Y. The analogues are used to treat

CC infections caused by bacteria (Gram positive or negative, or anaerobic);

CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or

CC trematodes) or viruses. Typical of very many pathogens that can be

CC controlled are *Leishmania*, *Trypanosoma*, *Ascaris lumbricoides*, *Fasciola*

CC *hepatica*, *Klebsiella pneumoniae*, *Bordetella pertussis*, *Staphylococcus*

CC *aureus*, *Listeria*, *Clostridium*, *rotavirus* and *papilloma virus*. Compounds

CC derived from the analogues may be used similarly; the compounds may

CC also be prepared from antibiotics or antitumour agents. The analogues

CC may be used therapeutically or to coat medical devices; also they are

CC useful as surface disinfectants, as additives to shampoo or soaps, as

CC insecticides or herbicides, or as preservatives for foods and technical

CC materials. The analogues are administered by injection, lavage, orally

CC or topically, generally at 0.1-50 mg/Kg. These analogues have a broader

CC spectrum of activity than indolicidin and modification as compounds

CC reduces their toxicity.

XX

SQ Sequence 21 AA;

Query Match 41.1%; Score 99; DB 19; Length 21;

Best Local Similarity 64.3%; Pred. No. 5.9e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY 1 ILKKPMPWMPRRKHEAPEAPIMILKK 28

Db 1 ILKKMPWMPRRK-----MILKK 18

RESULT 7

AA291806

ID AAY91806 standard; Peptide; 21 AA.

XX AAY91806;

AC 06-JUN-2000 (first entry)

XX

DT Amino acid sequence of cationic peptide MBI 11D4CN.

XX

DE Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;

XX leukaemia; polyoxalkylene-modified; APO; lymphoma; multiple myeloma;

KW breast; lung; ovary; uterus; skin; prostate; liver; colon;

KW multidrug resistance.

XX

OS Synthetic.

XX

XX WO965506-A2.

PN

XX

PD 23-DEC-1999.

XX

PF 14-JUN-1999; 99WO-CA00552.

XX

PR 12-JUN-1998; 98US-0096541.

XX

PA (MICR-) MICROLOGIX BIOTECH INC.

XX

PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;

XX

DR MPI; 2000-223549/19.

PT Novel pharmaceutical composition containing optionally activated
 PT polyoxalkylene-modified cationic peptides, useful for treating tumours
 XX
 PS Disclosure: Page 15; 94pp; English.

CC This sequence represents a cationic peptide amino acid sequence, which
 CC can be used in the pharmaceutical composition of the invention. The
 CC invention relates to a pharmaceutical composition containing at least one
 CC activated polyoxalkylene (APO)-modified cationic peptide. The
 CC modification of peptides with APO increases their activity against tumour
 CC cells, including those with a multidrug resistant phenotype. The
 CC pharmaceutical composition can be used to treat tumours, specifically
 CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
 CC cervix, uterus, skin, prostate, liver and colon.

XX
 SQ Sequence 21 AA;

Query Match 41.1%; Score 99; DB 21; Length 21;
 Best Local Similarity 64.3%; Pred. No. 5.9e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY 1 ILKKPMPWPKRKHAEPEAPIMILKK 28
 Db 1 ILKKPMPWPKR-----IMILKK 18

RESULT 8
 AAY24571
 ID AAY24571 standard; peptide; 21 AA.
 XX
 AC AAY24571;
 XX
 DT 18-AUG-1999 (first entry)
 XX
 DE Indolicidin analogue #23.
 XX
 KW Indolicidin: bacterial infection; photo-oxidised solubiliser;
 KM anticicrobial; antibiotic; antidiarrhythmic; surface disinfectant;
 KM additive; shampoo; soap; insecticide; herbicide; preservative;
 KM food; technical material.
 XX
 OS Synthetic.
 XX
 PN WO9807745-A2.
 XX
 PD 26-FEB-1998.
 XX
 PF 21-AUG-1997; 97WO-US14779.
 XX
 PR 13-JAN-1997; 97US-0034949.
 PR 21-AUG-1996; 96US-0024754.
 XX
 PA (MICR-) MICROLOGIX BIOTECH INC.
 XX
 PI Erfile D, Fraser JR, Krieger TJ, Taylor R, West MH;
 XX
 DR WPI; 1998-169090/15.
 XX
 PT New indolicidin analogues with antimicrobial activity and related
 PT nucleic acid - vectors, transformed cells and antibodies, also
 PT conjugates with polyoxalkylene glycol and fatty acid to reduce
 PT toxicity, useful therapeutically, as disinfectants etc.
 XX
 PS Claim 12; Page 89; 129pp; English.

CC AAY24549 to AAY24615 represent indolicidin analogues of formulae
 CC (I)-(VIII) containing up to 25 amino acids (aa): R₁XX₁XB (I). B₁XX₁XB
 CC (II). B₁XX₁XX₁XB (III). B₁XX₁XX₁BB₁(AA)MI₁BB₁GS (IV). B₁XX₁XX₁BB₁(AA)nm
 CC (V). LBB₁XX₁XX₁BB₁ (VI). LK₁XX₁XX₁BB₁ (VII) and BB₁XX₁XX₁BB₁ (VIII).
 CC where Z = F or V; X = hydrophobic residue, preferably W; B = basic aa,
 CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
 CC in (VIII) at least 2 X = F or Y. The analogues are used to treat

CC infections caused by bacteria (Gram positive or negative, or anaerobic);
 CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
 CC trematodes) or viruses. Typical of very many pathogens that can be
 CC controlled are Leishmania, trypanosoma, Ascatis tumoricoides, Fasciola
 CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
 CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
 CC derived from the analogues may be used similarly; the compounds may
 CC also be prepared from antibiotics or antidiarrhythmic agents. The analogues
 CC may be used therapeutically or to coat medical devices; also they are
 CC useful as surface disinfectants, as additives to shampoo or soaps, as
 CC insecticides or herbicides, or as preservatives for foods and technical
 CC materials. The analogues are administered by injection, lavage, orally
 CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
 CC spectrum of activity than indolicidin and modification as compounds
 CC reduces their toxicity.

XX
 SQ Sequence 21 AA;

Query Match 40.7%; Score 98; DB 19; Length 21;
 Best Local Similarity 64.3%; Pred. No. 7.8e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY 1 ILKKPMPWPKRKHAEPEAPIMILKK 28
 Db 1 ILKKPMPWPKR-----IMILKK 18

RESULT 9
 AAY91808
 ID AAY91808 standard; Peptide; 21 AA.
 XX
 AC AAY91808;
 XX
 DT 06-JUN-2000 (first entry)
 XX
 DE Amino acid sequence of cationic peptide MBI 11D6CN.
 XX
 KW Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
 KM leukemia; polyoxalkylene-modified; APO; lymphoma; multiple myeloma;
 KM breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
 KM multidrug resistance.
 XX
 OS Synthetic.
 XX
 PN WO965506-A2.
 XX
 PD 23-DEC-1999.
 XX
 PF 14-JUN-1999; 99WO-CA00552.
 XX
 PR 12-JUN-1998; 98US-0096541.
 XX
 PA (MICR-) MICROLOGIX BIOTECH INC.
 XX
 PI Friedland HD, Krieger TJ, Taylor R, Erfile D, Fraser JR, West MHP;
 XX
 DR WPI; 2000-223549/19.
 XX
 PT Novel pharmaceutical composition containing optionally activated
 PT polyoxalkylene-modified cationic peptides, useful for treating tumours
 XX
 PS Disclosure: Page 15; 94pp; English.

CC This sequence represents a cationic peptide amino acid sequence, which
 CC can be used in the pharmaceutical composition of the invention. The
 CC invention relates to a pharmaceutical composition containing at least one
 CC activated polyoxalkylene (APO)-modified cationic peptide. The
 CC modification of peptides with APO increases their activity against tumour
 CC cells, including those with a multidrug resistant phenotype. The
 CC pharmaceutical composition can be used to treat tumours, specifically
 CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
 CC cervix, uterus, skin, prostate, liver and colon.

XX Sequence 21 AA: 40.7%; Score 98; DB 21; Length 21;
 Query Match Best Local Similarity 64.3%; Pred. No. 7.8e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

OY 1 ILKKPMPWPRRKHEAPEAPEIMILKK 28
 |||||
 DB 1 ILKKPMPWPRR-----IMILKK 18

RESULT 10
 AAY24570
 ID AAY24570 standard; peptide; 20 AA.
 AC AAY24570;
 DT 18-AUG-1999 (first entry)
 DE Indolicidin analogue #22.
 XX Indolicidin; bacterial infection; photo-oxidised solubilisier;
 KW antimicrobial; antibiotic; antitarrhythmic; surface disinfectant;
 KW additive; shampoo; soap; insecticide; herbicide; preservative;
 KW food; technical material.
 OS Synthetic.
 XX WO9807745-A2.
 PD 26-FEB-1998.
 XX 21-AUG-1997; 97MO-US14779.
 PF 13-JUN-1997; 97US-0034949.
 PR 21-AUG-1996; 96US-0024754.
 XX (MICR-) MICROLOGIX BIOTECH INC.
 PA Effle D, Fraser JR, Krieger TJ, Taylor R, West MH;
 PI WPI; 1998-169090/15.
 DR New indolicidin analogues with antimicrobial activity and related
 PT nucleic acid - vectors, transformed cells and antibodies, also
 PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
 PT toxicity, useful therapeutically, as disinfectants etc.
 XX Claim 12; Page 89; 129pp; English.

XX AAY24549 to AAY24615 represent indolicidin analogues of formulae
 CC (1)-(VIII) containing up to 25 amino acids (aa): RXXXXB (I), BXXXXB
 CC (II), BBXXB (III), BXXXB (IV), BXXXB (V), BXXXB (VI),
 CC (V), LBXXB (VI), LKXXB (VII), LKXXB (VIII), LKXXB (IX),
 CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
 CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
 CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
 CC infections caused by bacteria (Gram positive or negative, or anaerobic);
 CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
 CC trematodes) or viruses; typical of very many pathogens that can be
 CC controlled are leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
 CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
 CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
 CC derived from the analogues may be used similarly; the compounds may
 CC also be prepared from antibiotics or antitarrhythmic agents. The analogues
 CC may be used therapeutically or to coat medical devices; also they are
 CC useful as surface disinfectants, as additives to shampoo or soaps, as
 CC insecticides or herbicides, or as preservatives for foods and technical
 CC materials. The analogues are administered by injection, lavage, orally
 CC or topically, generally at 0.1-50 mg/Kg. These analogues have a broader
 CC spectrum of activity than indolicidin and modification as compounds
 CC reduces their toxicity.

XX Sequence 20 AA: 38.8%; Score 93.5; DB 19; Length 20;
 Query Match Best Local Similarity 60.7%; Pred. No. 2.7e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

OY 1 ILKKPMPWPRRKHEAPEAPEIMILKK 28
 |||||
 DB 1 ILKKPMPWPRR-----MILKK 17

RESULT 11
 AAY91807
 ID AAY91807 standard; Peptide; 20 AA.
 AC AAY91807;
 DT 06-JUN-2000 (first entry)
 DE Amino acid sequence of cationic peptide MBI 11D5CN.
 XX Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
 KW leukaemia; polyoxyalkylene-modified; APO; lymphoma; multiple myeloma;
 KW breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
 KW multidrug resistance.
 OS Synthetic.
 XX WO965506-A2.
 PD 23-DEC-1999.
 XX 14-JUN-1999; 99MO-CA00552.
 PF 12-JUN-1998; 98US-0096541.
 PR (MICR-) MICROLOGIX BIOTECH INC.
 PA Friedland HD, Krieger TJ, Taylor R, Effle D, Fraser JR, West MHP;
 PI WPI; 2000-223549/19.
 DR Novel pharmaceutical composition containing optionally activated
 PT polyoxyalkylene-modified cationic peptides, useful for treating tumours
 PT
 PT Disclosure; Page 15; 94pp; English.

XX This sequence represents a cationic peptide amino acid sequence, which
 CC can be used in the pharmaceutical composition of the invention. The
 CC invention relates to a pharmaceutical composition containing at least one
 CC activated polyoxyalkylene (APO)-modified cationic peptide. The
 CC modification of peptides with APO increases their activity against tumour
 CC cells, including those with a multidrug resistant phenotype. The
 CC pharmaceutical composition can be used to treat tumours, specifically
 CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
 CC cervix, uterus, skin, prostate, liver and colon.

XX Sequence 20 AA: 38.8%; Score 93.5; DB 21; Length 20;
 Query Match Best Local Similarity 60.7%; Pred. No. 2.7e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 11; Gaps 1;

OY 1 ILKKPMPWPRRKHEAPEAPEIMILKK 28
 |||||
 DB 1 ILKKPMPWPRR-----MILKK 17

RESULT 12
 AAM12873
 ID AAM12873 standard; peptide; 13 AA.

XX AAM12873;
 AC 10-DEC-1997 (first entry)
 DT XX
 DE Antimicrobial cationic peptide CP-11.
 KW Bacterial; viral; antitumour; food; preservative; inhibitor; growth;
 KM bacterium; yeast; endotoxaemia; sepsis; antibiotic; fungal;
 KW antiviral; Candida albicans; steriliant; Salmonella; yersinia;
 KM Shigella.
 OS Synthetic.
 PN WO9708199-A2.
 PD 06-MAR-1997.
 PE 23-AUG-1996; 96WO-IB00996.
 PR 23-AUG-1995; 95US-0002687.
 XX (UYBR-) UNIV BRITISH COLUMBIA.
 PA Falla TJ, Gough M, Hancock RM;
 PI WPI; 1997-179179/16.
 DR Cationic peptide(s) having anti-microbial activity - used for the
 PT inhibition of bacterial and viral growth, as an antitumour agent,
 PT and as a food preservative
 PS Claim 2; Page 65; 89pp; English.
 XX The present sequence represents a specifically claimed novel isolated
 CC cationic peptide which has antimicrobial activity. The amino acid
 CC sequence of antimicrobial cationic peptides (including the present
 CC sequence) is selected from: X1X1ProX2X3X2Pro(X2X2Pro)(nX2X3(X5)O;
 CC X1X1ProX2X3X4(X5)ProX2X3X3; X1X1X3(ProTrp)uX3X2X5X2X5X2(X5)O;
 CC X1X1X3X3X2Pro(X2X2Pro)(nX2(X5)m: where m = 1-5; n = 1-2; O = 2-5; r
 CC = 0-8; u = 0-1; X1 = Ile, Leu, Val, Phe, Tyr, Trp or Met; X2 = Trp or
 CC Phe; X3 = Arg or Lys; X4 = Trp or Lys; and X5 = Phe, Trp, Arg, Lys or
 CC Pro. The peptides are preferably amidated or carboxymethylated. The
 CC peptides may be used in methods for inhibiting the growth of a bacterium
 CC or yeast, or for inhibiting an endotoxaemia or sepsis associated
 CC disorder in a subject. The peptides have a broad activity against
 CC antibiotic resistant bacteria, combined with activity against the
 CC medically important fungus Candida albicans. In addition, the peptides
 CC are useful as antitumour agents and/or antiviral agents. The peptides
 CC may be used as sterilants or preservatives of materials susceptible to
 CC microbial or viral contamination, e.g. in processed foods to inhibit
 CC Salmonella, Yersinia and Shigella. The peptides are compact and tend to
 CC have a unique polypyrrole type II extended helix structure that permits
 CC them to span the membrane with relatively few amino acids. The peptides
 CC possess the ability to work synergistically with antibiotics, and in
 CC addition, some of them possess anti-endotoxin activity.
 CC N.B. The present sequence represents SEQ ID NO:1 in the claims and
 CC examples of the specification, but differs slightly from the SEQ ID NO:1
 CC in the sequence listing on page 51 of the specification (see AAM27179).
 CC
 SQ Sequence 13 AA;

Query Match 37.8%; Score 91; DB 18; Length 13;
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPMPWRRK 13
 |||
 DB 1 ILKKPMPMPWRRK 13

RESULT 13
 AAM24609

ID AAY24609 standard; peptide; 13 AA.
 XX AAY24609;
 AC 18-AUG-1999 (first entry)
 DT XX
 DE Indolicidin analogue #61.
 KW Indolicidin; bacterial infection; photo-oxidised solubiliser;
 KM antimicrobial; antibiotic; antiarhythmic; surface disinfectant;
 KW additive; shampoo; soap; insecticide; herbicide; preservative;
 KM food; technical material.
 OS Synthetic.
 PN WO9807745-A2.
 PD 26-FEB-1998.
 PE 21-AUG-1997; 97WO-US14779.
 PR 13-JAN-1997; 97US-0034949.
 PR 21-AUG-1996; 96US-0024754.
 XX (MICR-) MICROLOGIX BIOTECH INC.
 PA Erffle D, Fraser JR, Krieger TJ, Taylor R, West MH;
 PI WPI; 1998-169090/15.
 DR New indolicidin analogues with antimicrobial activity and related
 PT nucleic acid - vectors, transformed cells and antibodies, also
 PT conjugates with polyoxyalkylene glycol and fatty acid to reduce
 PT toxicity, useful therapeutically, as disinfectants etc.
 PS Example 1; Page 32; 129pp; English.
 XX AAY24549 to AAY24615 represent indolicidin analogues of formulae
 CC (I)-(VIII) containing up to 25 amino acids (aa): RXXXXXB (I), BXXXXXB
 CC (II), BBRXXXXXB (III), BXXXXBBBn(AA)nMLBBAGS (IV), BXXXXBB(AA)nM
 CC (V), LBBnXXnXXnXR (VI), LKnXXXXRRR (VII) and BBRXXXXBBB (VIII).
 CC Where Z = P or V; X = hydrophobic residue, preferably W; B = basic aa,
 CC preferably R or K; AA = any aa; n = 0 or 1; in (II), at least 1 Z = V;
 CC in (VIII) at least 2 X = F or Y. The analogues are used to treat
 CC infections caused by bacteria (gram positive or negative, or anaerobic);
 CC fungi (yeast or moulds); parasites (protozoa, nematodes, cestodes or
 CC trematodes) or viruses. Typical of very many pathogens that can be
 CC controlled are Leishmania, Trypanosoma, Ascaris lumbricoides, Fasciola
 CC hepatica, Klebsiella pneumoniae, Bordetella pertussis, Staphylococcus
 CC aureus, Listeria, Clostridium, rotavirus and papilloma virus. Compounds
 CC derived from the analogues may be used similarly; the compounds may
 CC also be prepared from antibiotics or antiarhythmic agents. The analogues
 CC may be used therapeutically or to coat medical devices; also they are
 CC useful as surface disinfectants, as additives to shampoo or soaps, as
 CC insecticides or herbicides, or as preservatives for foods and technical
 CC materials. The analogues are administered by injection, lavage, orally
 CC or topically, generally at 0.1-50 mg/kg. These analogues have a broader
 CC spectrum of activity than indolicidin and modification as compounds
 CC reduces their toxicity.
 CC
 SQ Sequence 13 AA;

Query Match 37.8%; Score 91; DB 19; Length 13;
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPMPWRRK 13
 |||
 DB 1 ILKKPMPMPWRRK 13

RESULT 14
 AAM66378

```

ID AAM6378 standard; peptide; 13 AA.
XX
AC AAM6378;
XX
DT 12-JAN-1999 (first entry)
XX
DE Cationic peptide of claim 15 #5.
XX
KW Indolicidin analogue; resistance; cationic peptide; antibiotic;
KW bacterial infection; tolerance; antibacterial; microorganism;
KW bacteria; fungus; parasite; virus.
XX
OS Synthetic.
XX
PN WO9840401-A2.
XX
PD 17-SEP-1998.
XX
PF 10-MAR-1998; 98WO-CA00190.
XX
PR 25-FEB-1998; 98US-0030619.
PR 10-MAR-1997; 97US-0040649.
PR 20-AUG-1997; 97US-0915314.
PR 26-SEP-1997; 97US-0060099.
XX
PA (MICR-) MICROLOGIX BIOTECH INC.
XX
PI Fraser JR, McNicol PJ, West MHP;
XX
DR WPI; 1998-520800/44.
XX
PT New indolicidin peptide analogues - useful for, e.g. enhancing
PT activity of antibiotic or overcoming tolerance, acquired resistance
PT or inherent resistance of microorganisms
XX
PS Claim 15; Page 93; 105pp; English.
XX
CC The present sequence represents a specifically claimed cationic peptide
CC from the present invention. The present invention describes compositions
CC and methods for treating infection, especially bacterial infections. The
CC compositions and methods use cationic peptides in combination with an
CC antibiotic agent which are then administered to a patient to enhance the
CC activity of the antibiotic agent, to overcome: (a) tolerance; (b)
CC acquired resistance; and (c) inherent resistance. The combinations of
CC antibiotics and cationic peptides can provide synergistic activity
CC against a microorganism that is tolerant, inherently resistant, or has
CC acquired resistance to an antibiotic agent. They can be used for killing
CC e.g. bacteria, fungi, parasites and viruses.
XX
SQ Sequence 13 AA;

Query Match 37.8%; Score 91; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWPRRK 13
   |||||
DB 1 ILKKPMPWPRRK 13

RESULT 15
AAM71690
ID AAM71690 standard; Peptide; 13 AA.
XX
AC AAM71690;
XX
DT 11-JAN-1999 (first entry)
XX
DE Cationic peptide MB111 (MW 1879).
XX
KW MB111; cationic peptide; plasmid pK1; small cryptic plasmid;
KW replication; Repa; vector; RAMP.
XX

```

```

OS Synthetic.
XX
PN WO9841636-A2.
XX
PD 24-SEP-1998.
XX
PF 16-MAR-1998; 98WO-CA00214.
XX
PR 14-MAR-1997; 97US-0040722.
XX
PA (BURI/) BURIAN J.
PA (KAYW/) KAY W W.
XX
PI Burian J, Kay WW;
XX
DR WPI; 1998-531571/45.
XX
PT Increasing plasmid copy number in a cell with the repa gene product
PT - and an small cryptic plasmid ori sequence, useful for high level
PT expression of e.g. cytokines, antigens or therapeutic proteins
XX
PS Example 13; Page 54; 82pp; English.
XX
XX MB111 is a small (mol.wt. 1879) cationic peptide. DNA encoding
CC MB111 has been incorporated into vector pR2h-B1, in which the
CC replication leader (R21) sequence of Repa (see also AAM71686) is
CC joined to 2 Hpro peptides (see also AAM71692), to provide a
CC vector for expression of MB111 in host cells. The invention
CC provides controlled replication plasmid vectors (RAMP vectors)
CC comprising a replicated replication origin of a small cryptic plasmid and a
CC gene encoding Repa. The vectors can reach very high levels of
CC plasmid replication, but are not lethal to the host cell, and can
CC be used to direct the high level expression of e.g. cytokines,
CC antigens and therapeutic proteins.
XX
SQ Sequence 13 AA;

Query Match 37.8%; Score 91; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILKKPMPWPRRK 13
   |||||
DB 1 ILKKPMPWPRRK 13

Search completed: January 15, 2003, 18:08:34
Job time : 36 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2002, 15:37:24 : Search time 28 seconds
(without alignments)
272.277 Million cell updates/sec

Title: US-09-444-281-35-27-35
Perfect score: 241
Sequence: 1 ILKKWPMWRKHEAPEAPETMLKKWPMWRK 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.todent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriophage:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.5	29.3	1245	3	Q9Y7V5
2	67.5	28.0	746	12	Q9JH31
3	67	27.8	723	12	Q9DUC4
4	64	26.6	175	12	Q91R08
5	63	26.1	49	12	Q9D780
6	63	26.1	192	16	Q928B7
7	63	26.1	748	12	Q9D781
8	63	26.1	750	12	Q91D04
9	62	25.7	367	11	Q63778
10	61	25.3	92	12	Q8V7E2
11	61	25.3	250	17	Q8TWC0
12	61	25.3	284	10	Q94CT8
13	61	25.3	734	12	Q8V711
14	61	25.3	735	12	Q9DUC9
15	61	25.3	985	15	Q98414
16	60	24.9	147	11	Q61427

17	60	24.9	242	17	Q8TVM8
18	60	24.9	381	16	Q9A7E1
19	60	24.9	766	12	Q91CY5
20	59.5	24.7	114	16	Q9X8C2
21	59.5	24.7	2292	12	Q66765
22	59	24.5	95	10	Q9LON0
23	59	24.5	485	16	P72844
24	59	24.5	739	12	Q9PAQ3
25	58.5	24.3	521	10	Q94EF3
26	58.5	24.3	1367	2	Q9ADM1
27	58	24.1	107	16	Q9XAE4
28	58	24.1	513	11	Q63289
29	58	24.1	646	11	Q63779
30	58	24.1	879	11	Q8Y199
31	58	24.1	1300	11	P97692
32	57.5	23.9	117	15	Q9YR99
33	57.5	23.9	341	13	Q94516
34	57.5	23.9	376	3	Q94516
35	57.5	23.9	444	4	Q9HC40
36	57.5	23.9	772	4	Q9BXV6
37	57	23.7	252	15	Q9TUC7
38	57	23.7	426	12	Q99A07
39	57	23.7	970	11	Q88821
40	57	23.7	971	11	Q70458
41	56.5	23.4	157	5	Q9YOE8
42	56.5	23.4	162	5	Q9W1W7
43	56.5	23.4	532	16	Q9CK19
44	56	23.2	252	15	Q91TX8
45	56	23.2	252	15	Q91TV2

ALIGNMENTS

RESULT 1
Q9Y7V5 PRELIMINARY: PRT: 1245 AA.
AC Q9Y7V5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Candidospor surface protein.
GN CMPL.
GN Trichoderma harzianum.
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Mitosporic Hypocreales; Trichoderma.
OX NCBI_TaxID=5544;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 32173;
RX MEDLINE=9343861; PubMed=10413618;
RA Puyesky M., Benhamou N., Ponce Noyola P., Bauw G., Ziv T.,
Van Montagu M., Herrera Estrella A., Horwitz B.A.;
RT "Developmental regulation of cmpl, a gene encoding a multidomain
RT Candidospor surface protein of Trichoderma.";
RL Fungal Genet. Biol. 27:88-99(1999).
DR EMBL: AJ133651; CAB40845.1;
DR HSSP: P01180; INPO.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR001673; S_mold_repeat.
DR ProDom: PD006869; S_mold_repeat.2.
DR PROSITE: PS01346; CLAUDIN; UNKNOWN.1.
SQ SEQUENCE 1245 AA: 135824 MW: 32496749AFA0CDF8 CRC64;

Query Match Best local Similarity 29.0%: Score 70.5; DB 3; Length 1245;
Matches 9; Conservative 2; Mismatches 9; Indels 11; Gaps 1;

QY 4 KWPMPWRKHEAPEAPETMLKKWPMWRK 34
DB 1185 RMOQMSWPRRG-----CWQMSM 1204

RESULT 2

Q9JH31 ID 09JH31 PRELIMINARY; PRT; 746 AA.
 AC Q9JH31;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ORF1.
 OS TT virus.
 OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JUN02;
 RA Okamoto H.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 OS SEQUENCE FROM N.A.
 RC STRAIN=JUN02;
 RX MEDLINE=20456801; PubMed=11003468;
 RA Ukita M., Okamoto H., Nishizawa T., Tawara A., Takahashi M.,
 R Iizuka M., Miyakawa Y., Mayumi M.;
 RT "The entire nucleotide sequences of two distinct TT virus (TTV)
 RT isolates (JUN01 and JUN02) remotely related to the original TTV
 RT isolates."
 RL Arch. Virol. 145:1543-1559(2000).
 DR EMBL; AB028669; BAA94878.1;
 DR InterPro: IPR004219; TTVirus_Unk.
 DR Pfam: PF02956; TT_ORF1.1.
 SQ SEQUENCE 746 AA; 86561 MW; E0B22953AE764E3E CRC64;

Query Match

Best Local Similarity 28.0%; Score 67.5; DB 12; Length 746;
 Matches 11; Conservative 2; Mismatches 5; Indels 15; Gaps 1;

OY 5 WPMWPMRRKHEAPEAPEDIMILKKMPMPWRRK 37
 | | | | |
 DB 3 WGMWRMR-----RMPARRRRR 20

RESULT 3

Q9DUC4 ID 09DUC4 PRELIMINARY; PRT; 723 AA.
 AC Q9DUC4;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ORF1.
 OS TT virus.
 OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MF-TTV9;
 RA Okamoto H.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 OS SEQUENCE FROM N.A.
 RC STRAIN=MF-TTV9;
 RX MEDLINE=20534983; PubMed=11080484;
 RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,
 RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
 RT "Species-specific TT viruses in humans and nonhuman primates and their
 RT phylogenetic relatedness."
 RL Virology 277:368-378(2000).
 DR EMBL; AB041959; BAB19313.1;
 DR InterPro: IPR001563; Serine_carpept.
 DR InterPro: IPR004219; TTVirus_Unk.
 DR Pfam: PF02956; TT_ORF1.1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; UNKNOWN.1.
 SQ SEQUENCE 723 AA; 85393 MW; 232D003098766344 CRC64;

Query Match 27.8%; Score 67; DB 12; Length 723;

Best Local Similarity 34.4%; Pred. No. 3.6;
 Matches 11; Conservative 1; Mismatches 2; Indels 18; Gaps 1;

OY 6 WPMWPMRRKHEAPEAPEDIMILKKMPMPWRRK 37
 | | | | |
 DB 2 PMWPMRR-----WRRRRR 15

RESULT 4

Q91RD8 ID 091RD8 PRELIMINARY; PRT; 175 AA.
 AC Q91RD8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ORF3.
 OS TT virus.
 OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L03;
 RA Liu Z.H., Luo K.X., Hu J., He H.T.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF371370; AAK54733.1;
 DR InterPro: IPR004219; TTVirus_Unk.
 DR Pfam: PF02956; TT_ORF1.1.
 SQ SEQUENCE 175 AA; 22073 MW; 5212D7DA3FD72F81 CRC64;

Query Match

Best Local Similarity 26.6%; Score 64; DB 12; Length 175;
 Matches 11; Conservative 2; Mismatches 4; Indels 16; Gaps 2;

OY 5 WPMWPMRRKHEAPEAPEDIMILKKMPMPWRRK 37
 | | | | |
 DB 3 WSMW-WRRRR-----WMPRRR 19

RESULT 5

Q9DTR0 ID 09DTR0 PRELIMINARY; PRT; 49 AA.
 AC Q9DTR0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ORF1 (Fragment).
 OS TT virus.
 OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TYM9;
 RX MEDLINE=20566739; PubMed=11118348;
 RA Okamoto H., Nishizawa T., Tawara A., Takahashi M., Kishimoto J.,
 RA Sai T., Sugai Y.;
 RT "TT virus mRNAs detected in the bone marrow cells from an infected
 RT individual."
 RL Biochem. Biophys. Res. Commun. 279:700-707(2000).
 DR EMBL; AB050449; BAB19930.1;
 FT NON_TER
 SQ SEQUENCE 49 AA; 7225 MW; 1DA6F8F1AB69AA43 CRC64;

Query Match 26.1%; Score 63; DB 12; Length 49;
 Best Local Similarity 30.3%; Pred. No. 0.79;
 Matches 10; Conservative 3; Mismatches 4; Indels 16; Gaps 2;

OY 5 WPMWPMRRKHEAPEAPEDIMILKKMPMPWRRK 37
 | | | | |
 DB 3 WTW-WRRRR-----WPMRRR 19

RESULT 6

Q9ZBB7

```

RT individual.",
RL Biochem. Biophys. Res. Commun. 279:700-707(2000).
DR EMBL; AB050448; BAB19928.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
Pfam; PF02956; TT_ORF1.1.
SQ SEQUENCE 748 AA; 88552 MW; D65CCB2CA5CE26F CRC64;

Query Match          26.1%; Score 63; DB 12; Length 748;
Best Local Similarity 30.3%; Pred. No. 11;
Matches 10; Conservative 3; Mismatches 4; Indels 16; Gaps 2;

QY      5 WPMWPMRRKHAEPEAPRIMILKKMPMWRK 37
        ||| | : :
DB       3 WTWW-WQRRRR-----WPWRRR 19

RESULT 8
O91D04
AC O91D04 PRELIMINARY; PRT; 750 AA.
ID O91D04
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_Taxid=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21488921; PubMed=11601907;
RA Okamoto H., Nishizawa T., Takahashi M., Asabe S., Tsuda F.,
RA Yoshikawa A.;
RT "Heterogeneous distribution of TT virus of distinct genotypes in
RT multiple tissues from infected humans.";
RL Virology 288:358-368(2001).
DR EMBL; AB060592; BAB69900.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1.1.
SQ SEQUENCE 750 AA; 89223 MW; 616EC86DC3469091 CRC64;

Query Match          26.1%; Score 63; DB 12; Length 750;
Best Local Similarity 30.3%; Pred. No. 11;
Matches 10; Conservative 3; Mismatches 4; Indels 16; Gaps 2;

QY      5 WPMWPMRRKHAEPEAPRIMILKKMPMWRK 37
        ||| | : :
DB       3 WTWW-WQRRRR-----WPWRRR 19

RESULT 9
O63778
AC O63778 PRELIMINARY; PRT; 367 AA.
ID O63778
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Hypothetical 43.7 kDa protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE=87064324; PubMed=3023845;
RA D'Ambrosio E., Waltskin S.D., Witney F.R., Saleme A., Furano A.V.;
RT "Structure of the highly repeated, long interspersed DNA family (LINE
RT or L1RN) of the rat.";
RL Mol. Cell. Biol. 6:411-424(1986).
DR EMBL; M13100; AAA6046.1; -.
DR InterPro; IPR000566; Lipoclin_cyCFABP.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein.
```

SO SEQUENCE 367 AA; 43685 MW; 4654499CC185EC3B CRC64;

Query Match 25.7%; Score 62; DB 11; Length 367;

Best Local Similarity 39.4%; Pred. No. 7.4;

Matches 13; Conservative 4; Mismatches 12; Indels 4; Gaps 2;

OY 1 ILKMPWMPWR---RKHEAPEAPEIMILK-KW 29
| | | | | : | : | | | | |
DB 25 IFSKMCWFNRATCRMRQIDPSLSPCTKLKSKW 57

RESULT 10

O8V7E2 PRELIMINARY; PRT; 92 AA.

AC O8V7E2;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ORF1 (Fragment).

OS TT virus.

OC Viruses; ssDNA viruses; unclassified ssDNA viruses.

OX NCBI_TaxID=68887;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=21844401; PubMed=11855633;

RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,

RT Okamoto H.;

RT "Analysis of the complete genomes of thirteen TT virus variants

RT classifiable into the fourth and fifth genetic groups, isolated from

RT viremic infants."

RL Arch. Virol. 147:21-41(2002).

DR EMBL; AB064615; BAB79374.1; -.

DR InterPro: IPR004219; TVVirus_Unk.

DR Pfam: PF02956; TT_ORF1.1.

FT NON_TER 92

SO SEQUENCE 92 AA; 12429 MW; 188D883D05A7B09A CRC64;

Query Match 25.3%; Score 61; DB 12; Length 92;

Best Local Similarity 32.3%; Pred. No. 2.5;

Matches 10; Conservative 3; Mismatches 2; Indels 16; Gaps 1;

OY 7 WMPWRKHEAPEAPEIMILKMPWMPWRK 37
| | | | | : | : | | | | |
DB 3 WMPWRKHEAPEAPEIMILKMPWMPWRK 17

RESULT 11

O8TWG0 PRELIMINARY; PRT; 250 AA.

AC O8TWG0;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DE NAD-dependent protein deacetylase, STR2 family.

GN STR2 OR MK1075.

OS Methanopyrus kandleri.

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;

OX NCBI_TaxID=2320;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=AV19 / DSM 6324 / JCM 9639;

RA MEDLINE=21927647; PubMed=11930014;

RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,

RA Shchedrina O.V., Shakhova V.V., Belova G.I., Aravind L.,

RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,

RA Mal'kh A.G., Koonin E.V., Kozlovskiy S.A.;

RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19

RT and monophyly of archaeal methanogens."

RT Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).

DR EMBL; AE010396; AAM02288.1; -.

KW Complete proteome.

Query Match 25.3%; Score 61; DB 17; Length 250;

Best Local Similarity 50.0%; Pred. No. 6.7;

Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 3 KKMPWMPWRKHEAPEAPE 22
| | | | | : | : | | | | |
DB 60 KVMWILMRKRIAEAPNP 79

RESULT 12

O94C18 PRELIMINARY; PRT; 284 AA.

AC O94C18;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Glycine-rich protein IeGRP1.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;

RN [1]

RP SEQUENCE FROM N.A.

RA Uthappa M., Muchhal U.S., Baldwin J.C., Raghothama K.G.;

RT "IeGRP1: A new member of glycine-rich proteins from tomato

RT (Lycopersicon esculentum)."

RL Physiol. Plantarum 0:0-0(2001).

DR EMBL; AY026037; AAK08984.1; -.

SO SEQUENCE 284 AA; 23434 MW; E81A84C247CB9ED8 CRC64;

Query Match 25.3%; Score 61; DB 10; Length 284;

Best Local Similarity 31.2%; Pred. No. 7.6;

Matches 10; Conservative 0; Mismatches 4; Indels 18; Gaps 2;

OY 5 WMPWRKHEAPEAPEIMILKMPWMPWR 36
| | | | | : | : | | | | |
DB 9 WMPWRKHEAPEAPEIMILKMPWMPWR 22

RESULT 13

O8V7I1 PRELIMINARY; PRT; 734 AA.

AC O8V7I1;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ORF1.

OS TT virus.

OC Viruses; ssDNA viruses; unclassified ssDNA viruses.

OX NCBI_TaxID=68887;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=CT43F;

RA MEDLINE=21844401; PubMed=11855633;

RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,

RA Okamoto H.;

RT "Analysis of the complete genomes of thirteen TT virus variants

RT classifiable into the fourth and fifth genetic groups, isolated from

RT viremic infants."

RL Arch. Virol. 147:21-41(2002).

DR EMBL; AB064598; BAB79322.1; -.

DR InterPro: IPR004219; TVVirus_Unk.

DR Pfam: PF02956; TT_ORF1.1.

SO SEQUENCE 734 AA; 86978 MW; F60E188BC0104A68 CRC64;

Query Match 25.3%; Score 61; DB 12; Length 734;

Best Local Similarity 32.3%; Pred. No. 19;
Matches 10; Conservative 3; Mismatches 2; Indels 16; Gaps 1;

Qy 7 WPMRRKHEAPEEPIIMLKWPMPWRRK 37
| | : | | :
Db 3 WMYRRR-----PWRPWR 17

RESULT 14

09DUC9 PRELIMINARY; PRT; 735 AA.

AC 09DUC9; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)

OS ORF1

OC TT virus

OX NCBI_TaxID=68887;

RP SEQUENCE FROM N.A.

RA Okamoto H.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20534983; PubMed=11080484;

RA Okamoto H., Nishizawa T., Tawara A., Peng Y., Takahashi M.,

RT Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;

RT "Species-specific TT viruses in humans and nonhuman primates and their

phylogenetic relatedness.";

RL Virology 277:368-378(2000).

DR EMBL: AB041957; BAB19308.1;

DR InterPro: IPR004219; TTVirus_unk.

DR Pfam: PF02956; TT_ORF1; 1.

SQ SEQUENCE 735 AA; 86132 MW; 9ED818D6BE6FA5D3 CRC64;

Qy 5 WPMRRKHEAPEEPIIMLKWPMPWRRK 37
| | : | | :
Db 3 WPMRRRWRWRWR-----PWRWRWRWRWRWR 31

RESULT 15

098414 PRELIMINARY; PRT; 985 AA.

AC 098414; 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)

OS ENV.

OC Ovine lentivirus

OX NCBI_TaxID=11663;

RP SEQUENCE FROM N.A.

RX MEDLINE=95135990; PubMed=7834396;

RA Woodward T.M., Carlson J.O., de la Concha-Bermejo A.,

RT Demartini J.C.;

RT "Biological and genetic changes in ovine lentivirus strains following

passage in isogenic twin lambs.";

RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 8:124-133(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=85/34;

RA Carlson J.O., Demartini J.C., Mwaengo D.M.;

RT "Envelope glycoprotein nucleotide sequence and genetic

RT characterization of North American ovine lentiviruses.";

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U64439; AAB08725.1;

DR InterPro: IPR000328; Env_Gp41.

DR Pfam: PF00517; Gp41; 1.

KW Transmembrane.

SQ SEQUENCE 985 AA; 113794 MW; 3197258EDBE3597 CRC64;

Query Match 25.3%; Score 61; DB 15; Length 985;
Best Local Similarity 23.4%; Pred. No. 25;
Matches 18; Conservative 7; Mismatches 8; Indels 44; Gaps 5;

Qy 1 ILKK-----WPW-----WRRKHEAPEE----- 21
| | : | | :
Db 163 ILKRVYKQDPWNTYHWPMLQENMRQMKENERYGRITNKEDIDLLAKIRGRFC 222

Qy 22 --PIMILK--KMPWP 33
| : | | : | : |
Db 223 VPPFPALLCKTKKWCWP 239

Search completed: December 11, 2002, 15:39:39
Job time : 30 secs

THIS PAGE BLANK (USPTO)